

# Des métadonnées pour mutualiser les ressources

Archives ouvertes, OAI-PMH, portails ...

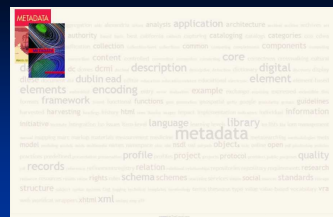
URFIST Bordeaux, 8 février 2011

Catherine Morel-Pair  
INIST-CNRS

[catherine.morel@inist.fr](mailto:catherine.morel@inist.fr)



[Licence Creative Commons](http://creativecommons.org/licenses/by/4.0/)



- Métadonnées ?
  - Généralités
  - Standards
    - Les langages d'encodage (HTML, XML)
    - Les standards de métadonnées
- L'interopérabilité en action : quelques exemples
  - Métadonnées et protocole OAI-PMH
    - Pour les archives ouvertes scientifiques et autres usages
  - Valorisation des thèses : projet TEF
  - Projet ORI-OAI
- La standardisation en action : quelques outils
- Conclusion- perspective
  - Métadonnées et "web de données" / web sémantique

15/02/2011

C. Morel-Pair

## I – Métadonnées ?

- Métadonnée : une donnée à propos d'une autre donnée
- Un jeu de métadonnées : un ensemble **structuré** d'information pour **“représenter”** des ressources
  - Ensemble de **propriétés/éléments** : titre, auteur, type ...
    - Renseignés par des “valeurs”
  - **Décrit, explicite, localise** la ressource
    - Pour en faciliter la recherche, l'usage, la gestion (NISO)
- Web et tous ensembles de ressources électroniques
  - Ou autres ...

15/02/2011

C. Morel-Pair

## Un monde numérique riche



15/02/2011

C. Morel-Pair

## ... des questions

... Des réponses :  
curation digitale intégrant des  
métadonnées standardisées



Haricots Bonduelle  
Contient du sodium  
Date limite 2010-10-10  
Distribué par X

15/02/2011

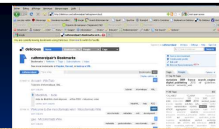
C. Morel-Pair



Resource discovery  
recherche, navigation

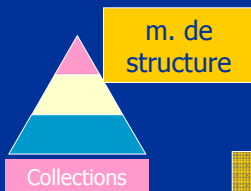


Labellisation  
Evaluation



Appropriation  
web 2.0 : tags, réseaux  
sociaux

m. descriptives du contenu



## Fonctions & typologie des métadonnées

m. administratives

m. de droits  
Propriété,  
responsabilité  
droits d'usage



m. techniques

Identification, localisation  
Caractéristiques fichiers



Interopérabilité  
mutualisation

Standards ...

Pérennisation

m. tous types  
m. techniques ++

15/02/2011

C. Morel-Pair

## A fonctions diverses standards et pratiques divers ...

- **Nombreux standards complémentaires**, parfois concurrents
  - Objectifs et objets décrits très divers
  - Acteurs divers
  - Complexité variable
- **Contenu des éléments** varié
  - Valeurs “texte”, contrôlées ou modélisées / texte libre
  - Valeurs complexes : formules, vignettes ...
- **Créées** manuellement / automatiquement / par transformation
- **Lisibles par l’homme et la machine**
- **Embarquées** ou **externes**

15/02/2011

C. Morel-Pair

## Embarquées ou externes ?

- “Embarquées”, fichier commun ressource-métadonnées
  - **Transportées avec la ressource**
    - en-tête (header) des documents **HTML**
    - Métadonnées images et multimédia natives
- Externes, dans un autre fichier
  - **Indispensable pour rechercher, gérer, échanger, mutualiser**
    - Catalogues, annuaires, notices bibliographiques
    - Archives ouvertes, protocole OAI-PMH et entrepôt de métadonnées
    - Métadonnées image ou multimedia **exportées** → BD, page Web

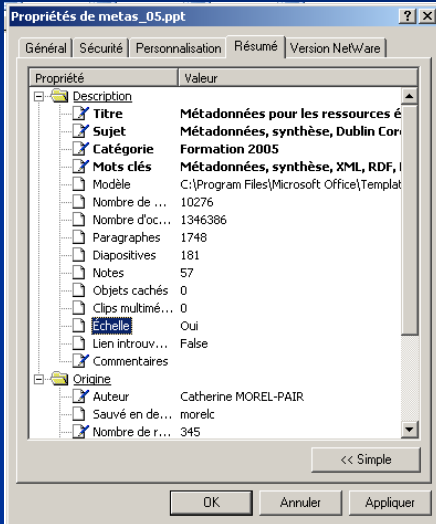
15/02/2011

C. Morel-Pair



# Des métadonnées “courantes”

## Métadonnées Windows



Explorateur Windows :  
clic droit sur un fichier  
“Propriétés”

! Variable entre versions  
Inconstantes lors de copies  
et autres migrations

“tags” images  
annuaires de sites  
cartes de sites  
Outils collaboratifs  
...

15/02/2011 C. Morel-Pair

! Variable entre versions  
Inconstantes lors de copies  
et autres migrations

- “tags” images
- annuaires de sites
- cartes de sites
- Outils collaboratifs
- ...

C. Morel-Pair

# Métadonnées image embarquées

## images

De formats propriétaires à des formats interopérables (XMP)

Description	Description
1	Le texte de présentation du projet de centre de compétences
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	

De formats propriétaires à des formats interopérables (XMP)

```
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:Description rdf:about=""
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  <dc:format>application/pdf</dc:format>
  <dc:title>
    <rdf:Alt>
      <rdf:li xml:lang="x-default">Le texte de présentation du projet de centre de comp
    </rdf:li>
    </dc:title>
  </dc:creator>
  <rdf:Seq>
    <rdf:li>Stéphane Pouyilau</rdf:li>
  </rdf:Seq>
  </dc:creator>
  <dc:subject>
    <rdf:Seq>
      <rdf:li>CNRS</rdf:li>
      <rdf:li>pôle de compétences</rdf:li>
      <rdf:li>images</rdf:li>
      <rdf:li>recherche scientifique</rdf:li>
      <rdf:li>SRM</rdf:li>
      <rdf:li>numération</rdf:li>
      <rdf:li>conservation</rdf:li>
      <rdf:li>standard</rdf:li>
      <rdf:li>seaux</rdf:li>
    </rdf:Seq>
  </dc:subject>
  </rdf:Description>
```

# Métadonnées images externes

## Photothèque collaborative Flickr'r

**[North Trondhjer (LOC)]**

**Notes:**  
Title from the Detroit Publishing Co., Catalogue J-foreign section, Detroit, Mich.: Detroit Publishing Company, 1905.  
Print no. 7107.  
Forms part of: Landscape and marine views of Norway in the Photochrom print collection.

**Subjects:**  
Norway--Trondheim.

**Format:** Photochrom prints--Color--1890-1900.

**Rights Info:** No known restrictions on publication.

**Repository:** Library of Congress, Prints and Photographs Division, Washington, D.C. 20540 USA, <http://loc.gov/loc/arp/arp.asp>

**Part Of:** Landscape and marine views of Norway (DLC) 2001695652

More information about the Photochrom Print Collection is available at <http://loc.gov/loc/arp/arp.asp>

**Persistent URL:** <http://loc.gov/loc/arp/arp.asp/06249>

**Call Number:** LOT 13422, no. 147 [Item]

Ces notes sont associées à cette photo. Cliquez le curseur sur la photo pour afficher les notes.

**Commentaires**

**Christian Seeneag** dit :  
View from the west-side of Trondheim, Byåsen.  
Ajouté il y a 2 mois. (permanente)

**lybion** dit :  
I belief the title should be: Trondheim (correct spelling) seen from west.  
Seen in your Photochrom Travel Views set. (V)  
Posté il y a 2 mois. (permanente)

**Christian Seeneag** dit :  
These titles are most likely the original ones. And at the time this photo was taken, the correct spelling was Trondhjem.  
Posté il y a 2 mois. (permanente)

**Souhaitez-vous faire un commentaire ?**  
[Créer votre compte arhalut](#) ou [connectez-vous](#) (si vous êtes déjà membre).

**Export MARC LOC →**

→ ... enrichissement catalogues LoC  
... ... interoperabilité

15/02/2011

## Plan de site

◆ Pour les professionnels		
◆ Actualités - Agenda		
<ul style="list-style-type: none"> <li>• <b>Dépôt légal</b> <ul style="list-style-type: none"> <li>- Qu'est-ce que le dépôt légal ?</li> <li>- Pourquoi déposer ?</li> <li>- Organisation</li> <li>- Législation</li> <li>- Chiffres-clés</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>- Tous les formulaires <ul style="list-style-type: none"> <li>• Bibliothèque nationale de France</li> <li>• Bibliothèques OL Imprimeur</li> <li>• Documents sonores</li> <li>• Documents multimédias, logiciels et bases de données</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>- Agenda <ul style="list-style-type: none"> <li>• Livres</li> <li>• Périodiques</li> <li>• Documents cartographiques</li> <li>• Musique imprimée</li> <li>• Vidéogrammes</li> <li>• Documents iconographiques</li> <li>• Sites Web</li> </ul> </li> </ul>
◆ ISSN, ISBN, autres numéros		
<ul style="list-style-type: none"> <li>- S'informer et obtenir un ISSN <ul style="list-style-type: none"> <li>• Qu'est-ce que l'ISSN ?</li> <li>• Pour quelles publications ?</li> <li>• Obligations de l'auteur</li> <li>• Obtenir un ISSN</li> <li>• Connaître l'ISSN d'une publication</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>- S'informer et obtenir un ISBN <ul style="list-style-type: none"> <li>• Qu'est-ce que l'ISBN ?</li> <li>• Pour quelles publications ?</li> <li>• Obtenir un ISBN</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>- S'informer sur les autres numérotations internationales <ul style="list-style-type: none"> <li>• DOI</li> <li>• EAIH</li> <li>• EIL</li> <li>• EINI</li> <li>• EIR</li> <li>• EIRN</li> <li>• EITC</li> <li>• EIVC</li> </ul> </li> </ul>
◆ Bibliographie nationale française		
<ul style="list-style-type: none"> <li>• <b>Catalogage et indexation</b> <ul style="list-style-type: none"> <li>- Description et accès <ul style="list-style-type: none"> <li>• Principes internationaux de catalogage <ul style="list-style-type: none"> <li>• Contexte</li> <li>• Les rencontres d'experts IME ICC</li> <li>• Les points forts</li> </ul> </li> <li>• Description bibliographique internationale normalisée (ISBD) <ul style="list-style-type: none"> <li>• Présentation</li> <li>• Texte de l'ISBD</li> </ul> </li> <li>• Resource Description and Access (RDA) <ul style="list-style-type: none"> <li>• Origines de RDA</li> <li>• Objectifs et élaboration de RDA</li> <li>• Caractéristiques de RDA</li> <li>• Mise en œuvre de RDA</li> <li>• RDA en France</li> </ul> </li> <li>• Normes françaises <ul style="list-style-type: none"> <li>• Recueil Normes de catalogage</li> </ul> </li> <li>• Description des manuscrits et fonds d'archives modernes</li> </ul> </li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>- Formats, encodage <ul style="list-style-type: none"> <li>• INTERMARC <ul style="list-style-type: none"> <li>• INTERMARC Bibliographique</li> <li>• Historique</li> </ul> </li> <li>• UNIMARC <ul style="list-style-type: none"> <li>• Notices bibliographiques</li> <li>• Nations d'origine</li> <li>• Données locales</li> <li>• Classifications</li> <li>• MARC en XML</li> <li>• MODS : Metadata Object Description Schema</li> <li>• Présentation du format</li> <li>• Principaux éléments</li> <li>• Exemples d'utilisation</li> </ul> </li> <li>• EAD : Encoded Archival Description</li> <li>• Dublin Core</li> <li>• METS : Metadata Encoding and Transmission Standard</li> <li>• MPEG-21 DIDL : Digital Item Declaration</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>- Protocoles d'échange de données <ul style="list-style-type: none"> <li>• Z39.50</li> <li>• OAI-PMH</li> <li>• SRU (Search/Retrieval via URL) <ul style="list-style-type: none"> <li>• Vue d'ensemble</li> <li>• Protocoles de transport</li> <li>• Formuler une recherche</li> </ul> </li> </ul> </li> <li>- Modélisation et ses applications <ul style="list-style-type: none"> <li>• Élaboration d'ontologies <ul style="list-style-type: none"> <li>• Modèles FRBR, FRAD et FRASAD</li> <li>• Modèle CIDOC CRM</li> <li>• Web sémantique, Web de données</li> </ul> </li> <li>• Présentation</li> <li>• Évaluation des catalogues</li> <li>• Projet BnF : data.bnf.fr</li> </ul> </li> <li>- Autorités <ul style="list-style-type: none"> <li>• Autorités BnF <ul style="list-style-type: none"> <li>• Présentation et statistiques</li> <li>• Personnes</li> <li>• Collectivités</li> <li>• Noms géographiques BnF</li> <li>• Titres uniformes</li> <li>• Marques commerciales</li> </ul> </li> <li>• RAMEAU <ul style="list-style-type: none"> <li>• Contours autorisés BnF, autorisés RAMEAU</li> </ul> </li> </ul> </li> </ul>

## Métadonnées et bibliothèques numériques

- “Organisations qui offrent des ressources, y compris en personnel, pour sélectionner, structurer, offrir un accès intellectuel, distribuer et conserver l'intégrité de documents sous une forme numérique” (IFLA, 1998)
- Tailles des collections, formats et types de documents, mode d'acquisition ... variés
  - Projet Gutenberg ... fonds anciens des BU... collections de revues ... Gallica ... Europeana : texte, image, multimedia
  - Objectifs patrimoine-culture et scientifique
- Catalogue et plus ...
  - Publication de ressources sur le web, droits, pérennité ...



## Les archives ouvertes scientifiques

- Ensembles de ressources scientifiques en libre accès
  - Interopérables via le protocole OAI-PMH de “moissonnage de métadonnées”
  - Ressources variées : articles, communications, thèses, cours, images, données de la recherche ...
  - Auto-soumission et “modération”
- Deux types
  - Institutionnelles : HAL, archives Universités ... ORI-OAI
  - Thématiques : PubMedCentral, RePec, ArXiv, E-LIS ...

15/02/2011

C. Morel-Pair

## Les évolutions du web

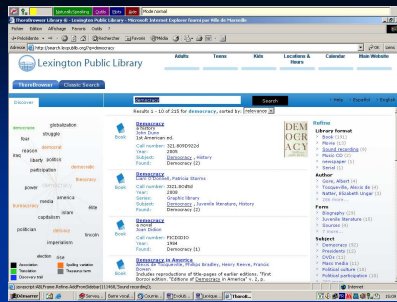


- Web sémantique : web de données
  - Granulaires : sujet-prédicat-objet ... “métadonnées” ?
  - Interprétables par les machines → structurées / standards (RDF, ontologies, URI)
  - Ouvertes et liées → nouvelles données
- Web 2.0 : web collaboratif ...
  - Ecriture simple et collaborative
  - Bookmarks et ressources partagées
  - Réseaux sociaux
  - Plateformes de services
  - Utilise “également” des standards
- ... “Web 3”

15/02/2011

C. Morel-Pair

# Evolutions des catalogues de bibliothèques



- Outil web
  - Interface
  - Ressources externes
- Documents plus complexes et liés
- Notices enrichies et "éclatées"
- Standards
  - métadonnées, BD et XML

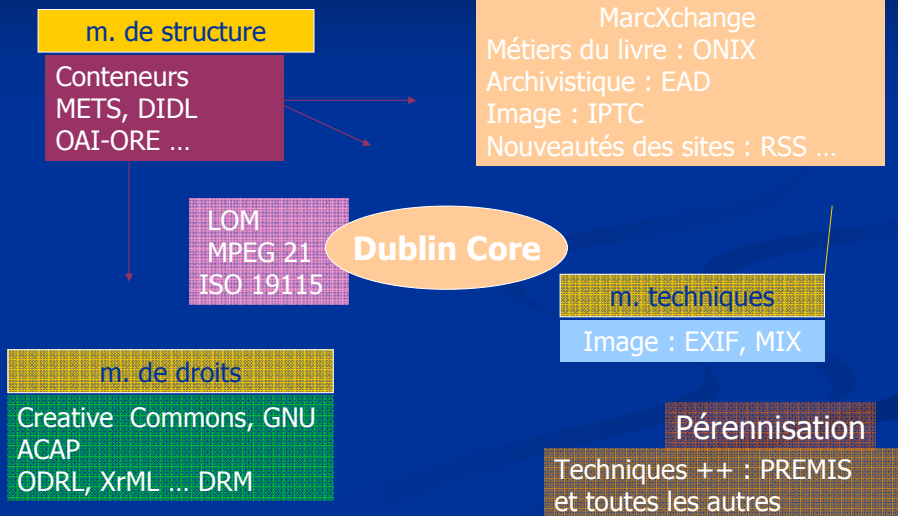
- Mutualisation, interopérabilité
  - Projets collaboratifs
- Interface de consultation ++
  - trier et affiner
  - moteur sémantique
  - clusters ...
- Nouveaux services
  - RSS, exports, Web 2.0...
- Demain : outil du web de données ?

Source : M. Gallon, BMVR Marseille, non publié

15/02/2011

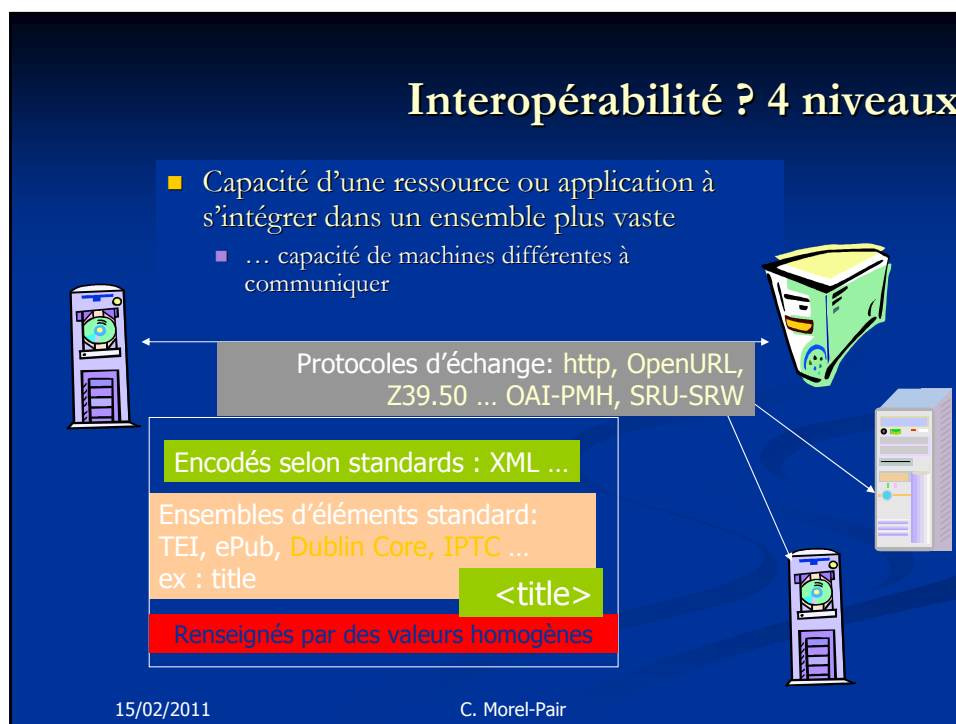
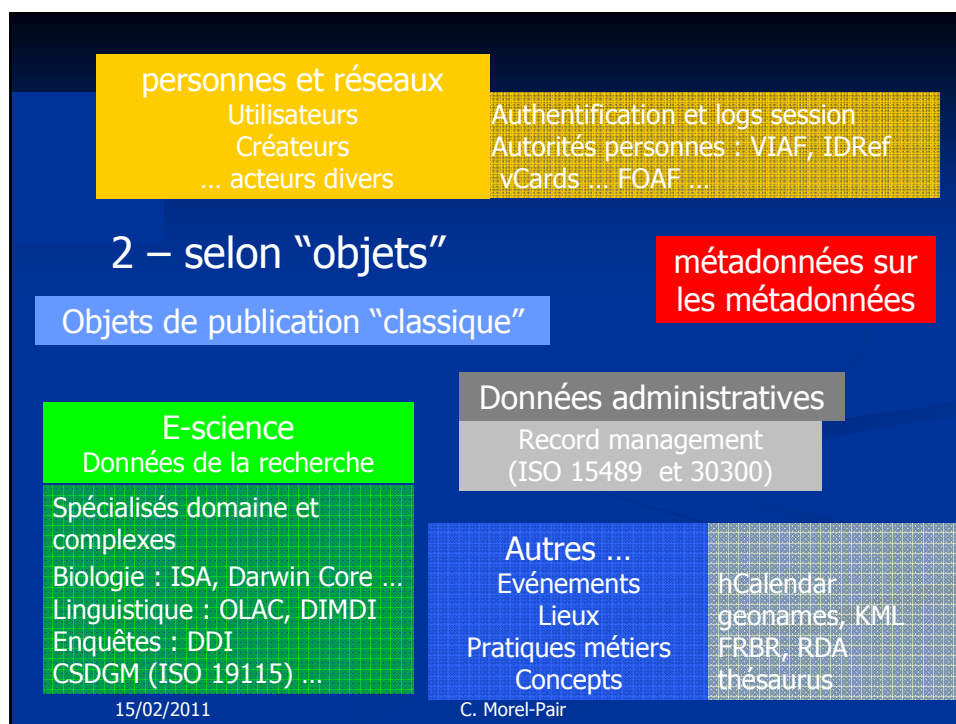
C. Morel-Pair

## Grands standards de métadonnées



15/02/2011

C. Morel-Pair



# Interopérabilité ? Un exemple : VIAF, virtual international authority file

The screenshot displays the VIAF website interface in French. At the top, the header reads 'VIAF' and 'Virtual International Authority File'. Below this, the 'Recherche' (Search) section shows the search criteria: 'Critère de recherche: mot(s) nom(s)' and 'Fichier source: Tout VIAF'. The search results are for 'Morelli, Jacopo, 1745-'. A list of 10 search results is shown, each with a title, a VIAF ID, and a link to the full record. The first result is 'Catálogo di commedie italiane [della raccolta di T. G. Farwell, compilato da J. Morelli] (2)'. Below the list, there is a 'Sélection de titres' (Selection of titles) section, a 'Pays de publication' (Country of publication) section with a world map, and a 'Statistiques par les décennies' (Statistics by decades) section. The date '15/02/2011' is visible in the bottom left corner.

[illegible]

**VIAF: Morelli, Jacopo, 1745-1819**  
 VIAF: Results of search: local.viafID exact "39367125"  
[Morelli, Jacopo, 1745-1819](#)  
 mardi 16 novembre 2010 07:11



## ■ Métadonnées ?

## ■ Des standards

### ■ Pour encoder du contenu : formats

#### ■ HTML, XML et les autres

### ■ Pour décrire : **métadonnées**

#### ■ **Dublin Core**, et une sélection

vocabulaires

15/02/2011


C. Morel-Pair

## Les standards du web dans Wikipedia

**Commentaires :**

The blog is written by Brian Kelly, UK Web Focus at URDLN, a national centre of expertise in digital information management based at Bath, U

**Formats de fichiers**



**Recent Posts**

- protocoles
- File Formats For Papers
- In
- Your Institutional Repository
- Geo-locating Your Event Tweets
- Amplified Events: Plans
- for
- Ha
- Th
- Pat
- Ac
- Ad
- It's Now Probably Time
- To Ditch Flash
- Web Accessibility - Code
- of Practice
- (BS 8878:2010)
- Evidence, Even If
- Flawed, For Blog Metrics
- One Minute Madness:
- Twitter: An Essential
- Tool
- for the Information Professional

**Recent Comments**

Jan Knowles on  
Amplified Events: Plans  
for...

As Lorcan Dempsey **suggested in a tweet** "The Wikipedia article traffic stats site mentioned in your comments is quite interesting, wonder how reliable it", I agree and thought I would explore what the statistics tell us about Wikipedia entries for a number of areas related to Web, metadata and related standards of interest to the JISC development community.

My survey was carried out on 6 July 2010. The following table provides a link to the relevant Wikipedia article, the data the article was created (with a link to the original page for the article), my comments on the article and the usage statistics for October 2009 and June 2010 (open to observe any significant variations).

Page	Created	Summary (subjective comments)	Stats: Oct 2009	Stats: Jun 2010
<a href="#">Linked Data</a>	May 2007	Multiple concerns have been identified with this article.	5,423	8,102
<a href="#">HTML</a>	Jul 2001	Appears to be a well-written and comprehensive article. Includes info box so factual information is available in DBPedia.	147,357	143,386
<a href="#">XML</a>	Sep 2001	Appears to be a well-written and comprehensive article. Includes info box so factual information is available in DBPedia.	159,749	126,599
<a href="#">XSLT</a>	Feb/Jun 2002	Appears to be a very thorough and comprehensive article. Includes info box so factual information is available in DBPedia.	7,160	18,938
<a href="#">RSS</a>	Sep 2002	Appears to be a well-written and comprehensive article. Includes a very brief info box so factual information is available in DBPedia.	7,160 (gaps)	18,938
<a href="#">AJAX (programming)</a>	Mar 2005	Appears factually correct.	98,629	90,300
<a href="#">SAML (Security Assertion Markup Language)</a>	Sep 2004	Appears to be a very thorough and comprehensive article.	4,421	2,942
<a href="#">Z39.50</a>	Oct 2004	Brief article which has been flagged as in need of improvements.	3,960	2,592
<a href="#">Search/Retrieve Web Service</a>	Feb 2004	Very little information provided.	506	462
<a href="#">Dublin Core</a>	Oct 2001	Appears factually correct though citations need improving.	7,013	7,501
<a href="#">METS (Metadata Encoding and Transmission Standard)</a>	Sep 2006	Appears factually correct though citations need improving.	7,236	4,573
<a href="#">MODS (Metadata Object Description Schema)</a>	Aug 2006	Appears to be a reasonable although succinct summary.	546	583

Extrapolating from the usage statistics for the two dates it would seem that popular articles such as HTML and XML have an annual number of views of around 1,750,000 and 1,720,000 whilst an article on a less well-known standard such as METS has an annual number of views of around 70,000. It is perhaps surprising, in light of the high viewing figures for METS that the annual viewing figures for MODS is around 6,700. Perhaps this is due to the name clash between the METS acronym and the Mets name used to refer to the New York Mets. However there isn't, as far as I am aware, such scope for confusion with names such as HTML, XML, SAML, etc.

What conclusion might we draw from such statistics? I would suggest that if I had an interest in ensuring that users had a good understanding of what Dublin Core is about and had access to the key sources of information then contributing to the Dublin Core Wikipedia page would be a good way of achieving that goal - after all the estimated viewing figures of around 87,000 surely can't be ignored.

Now Matt Jukes pointed out the potential difficulties of getting content into Wikipedia. But that is a question of "How we go about contributing to Wikipedia?" rather than "Should we?"

Can we accept that the answer to the second question should be "Yes" so that we can explore ways of addressing the first question?

★★★★★ Rate This


<http://ukwebfocus.wordpress.com/2010/07/08/how-well-read-are-technical-wikipedia-articles/>

subscribe to this bl  
receive notification  
posts by email.

**Blog Stats**

262,051 hits

Visitor location




ClustrMaps

**Syndicate This Page**

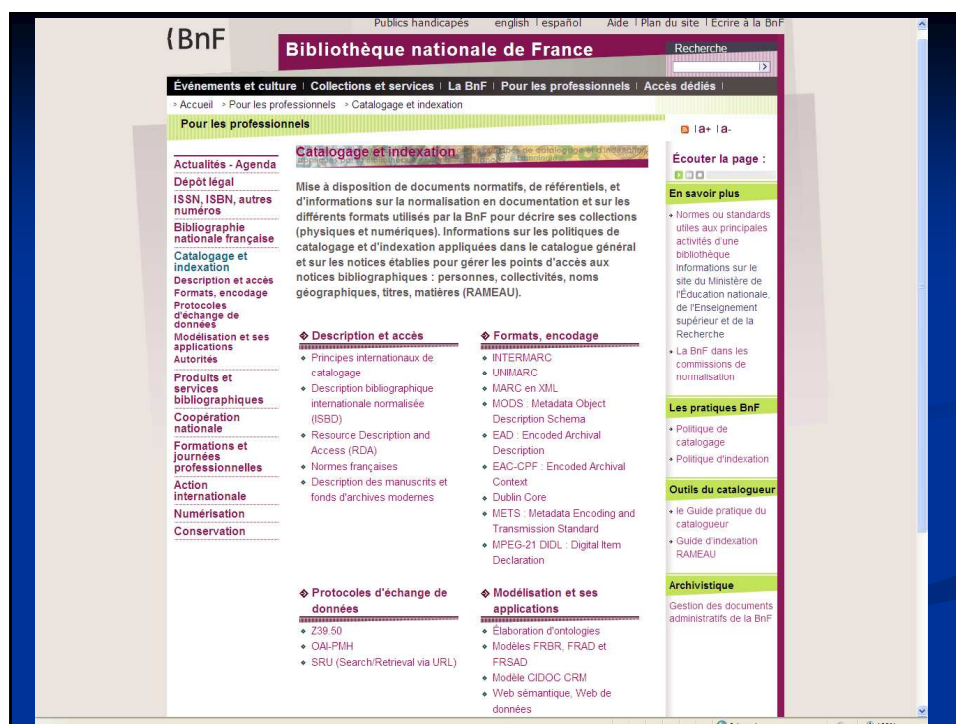
**RSS Feed for this**

**Licence**



This work is license  
a Creative Comm  
Attribution-  
Noncommercial-  
Allies 2.0 UK: Eng  
Wales License. It  
licence applies to b  
content published b  
author and (unless  
otherwise) guest b  
Comments posted  
blog will also be de  
have been publish  
licence. Please not  
that images and ot  
resources embedd  
blog may not be co  
such licences.

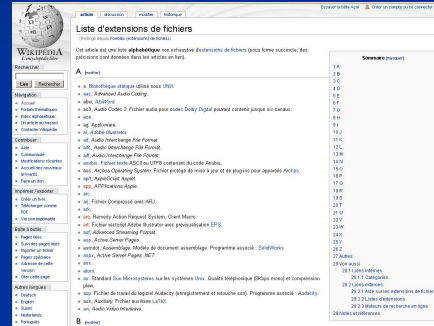
**About This Bl**



Plan du site					
Recherche dans le site					
Autres sites					
Marchés p					
L'ABES	Le réseau Sudoc	Recherche documentaire	Thèses	Manuscrits	IdRef - Autorités
Fournitu	Sudoc	document			
Formats d'encodage	<a href="#">URI</a>	URI resource pages / IFLA-CDNL (Conference of Directors of National Libraries) alliance for bibliographic standards.	2007	Standard international	
accès à l'information	<a href="#">RFC 2141</a> <a href="#">URN</a>	URN [Uniform Resource Name] syntax / IETF [Internet Engineering Task Force].	1997-	Standard international	
et gérer les droits de la propriété int	<a href="#">ARK</a>	ARK, Archival Resource Key [Ressource électronique] / California Digital Library. - 2007..	2007	Standard international	
nger des données bibliographiques	<a href="#">RFC 2426</a> <a href="#">VCard</a>	VCard / IETF [Internet Engineering Task Force].	1998	Standard international	
er, gérer et évaluer son activité	XML	Extensible Markup Language (XML) / World wide web consortium (W3C).		Standard international	
ver les documents	<a href="#">REC-xm1-names-19980114</a>	Namespaces in XML [Ressource électronique] / World Wide Web Consortium (W3C).	1999	Standard international	
ter une publication	<a href="#">RDF/XML Syntax Specification</a>	Resource description framework (RDF) [Ressource électronique] / World Wide Web Consortium (W3C).	1997-	Standard international	
l'Actu (Afnor CG46)	<a href="#">ISO 32000-1:2008</a> <a href="#">PDF</a>	Document management. Portable document format. Part 1, PDF 1.7.	2008	Norme internationale	
alphabets	<a href="#">RM</a>	métadonnées (RM) (disponible en anglais seulement).			
ctères de graphiques	<a href="#">RFC 2396</a> <a href="#">URI</a>	Uniform Resource Identifier (URI) : generic syntax / IETF [Internet Engineering Task Force].	1998	Standard international	

# Encodage des ressources Plus de 750 formats ...!

- Normes et standards, ou non
- Propriétaires
  - Outils d'édition et représentation ... propriétaires
  - Données "captives"
  - Ex : .doc, .xls, .ppt
- Libres - ouverts
  - Outils non limités
  - Souvent standards et basés sur des standards
  - Interopérables et pérennes
  - Ex : jpeg, langages de balisage, .odf



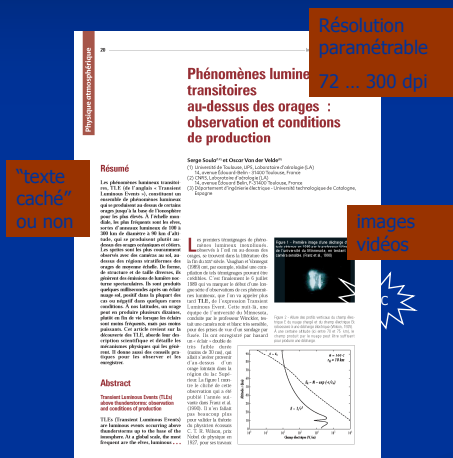
Quelques nuances ...

15/02/2011

C. Morel-Pair



# PDF, Portable Document Format



- Un des standards du Web
  - Usage massif
  - Lecteurs gratuits, toutes plateformes
  - Affichage fidèle des documents originaux
  - Contrôle des droits
- Format propriétaire
  - Versions successives, code et logiciels d'édition privés ...
- ... et ouvert
  - Indépendance matérielle et logicielle via des normes ISO
    - PDF/a (version 1.4), PDF/x
    - version 1.7 (2008)
- Métadonnées en XMP

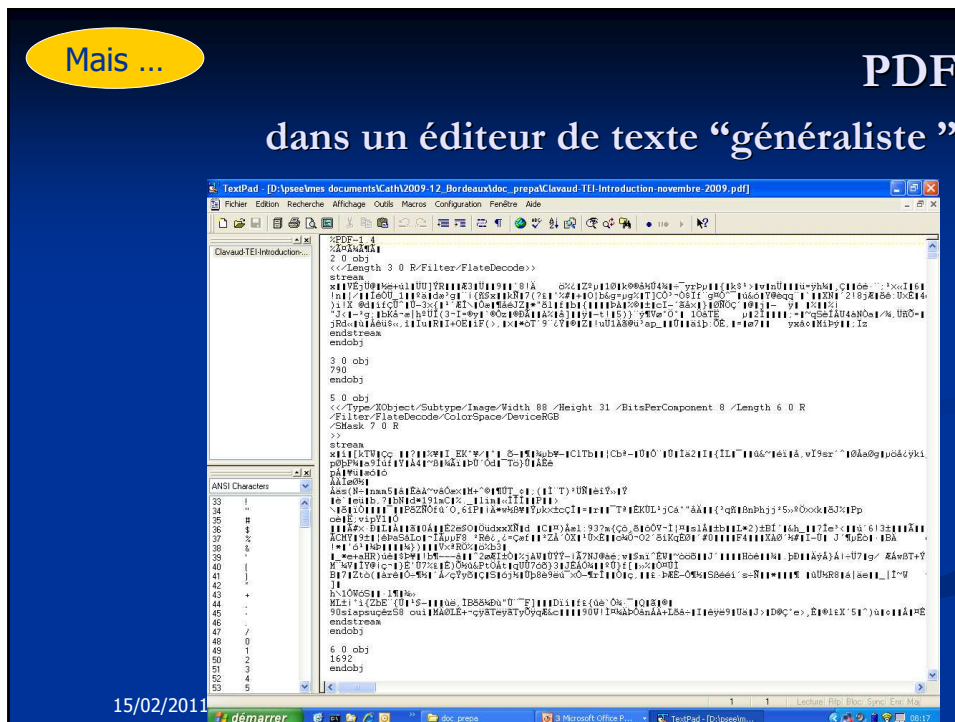
15/02/2011

C. Morel-Pair

Mais ...

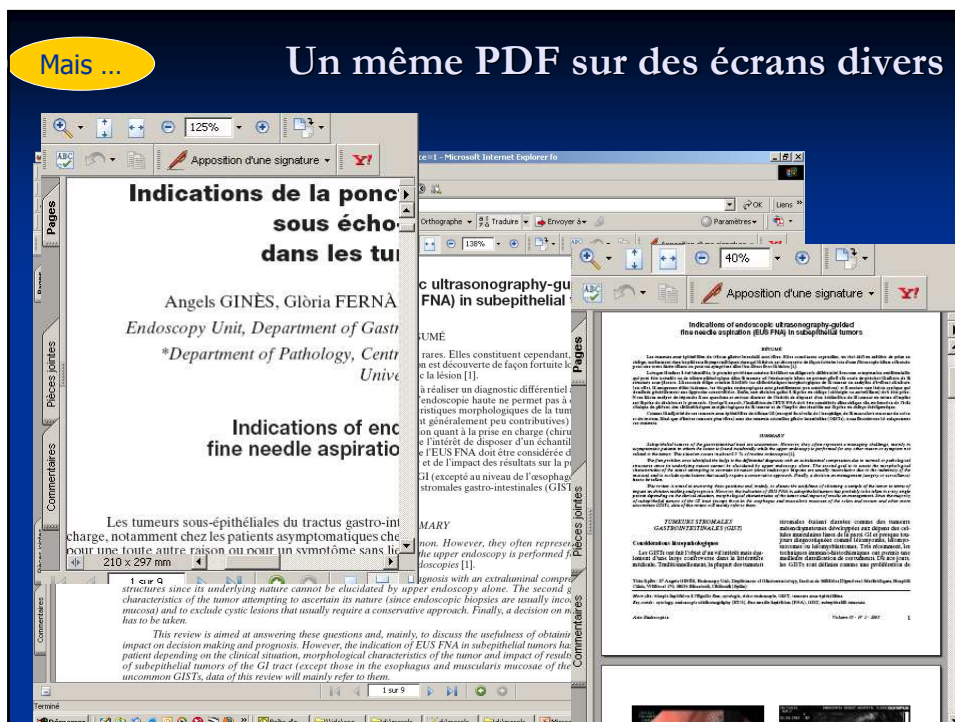
PDF

dans un éditeur de texte "généraliste"



Mais ...

Un même PDF sur des écrans divers



# Langages de balises

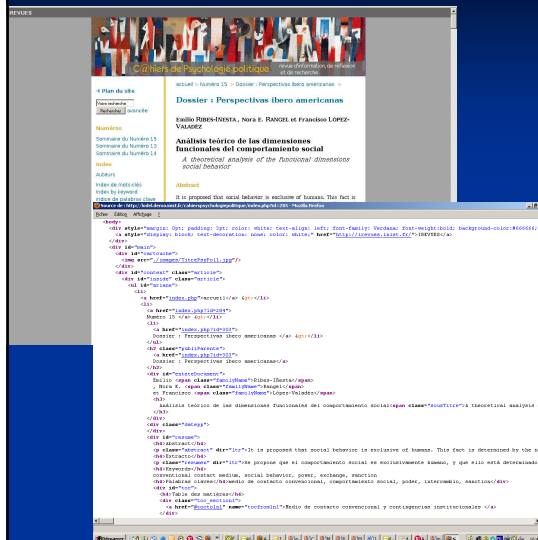
Des **standards** pour les ressources et les métadonnées

- Encodage des données selon une **structure logique**
  - Par des “**éléments**” matérialisés par des **<balises>**
    - Ex : **<title>XML et les métadonnées</title>**
  - Contenu manipulable par les machines
- **Dissocie le contenu de la présentation**
  - Des feuilles de style génèrent ensuite une ou des présentation(s)
  - Intermédiaire : HTML, langage de balises orienté présentation
- **Interopérabilité**
  - Ouverts, libres et standards
  - Usages multiples et pérennité des fichiers

15/02/2011

C. Morel-Pair

## 1 - HTML et XHTML



W3C Ubiquitous Web domain

- “LE” format du Web
- HyperText Mark-up Language, 1990
  - Document simple pour le Web
  - Liens hypertexte
  - Versions successives libres et ouvertes
  - XHTML : vers du vrai XML - écriture plus rigoureuse
- Interopérable avec applications diverses

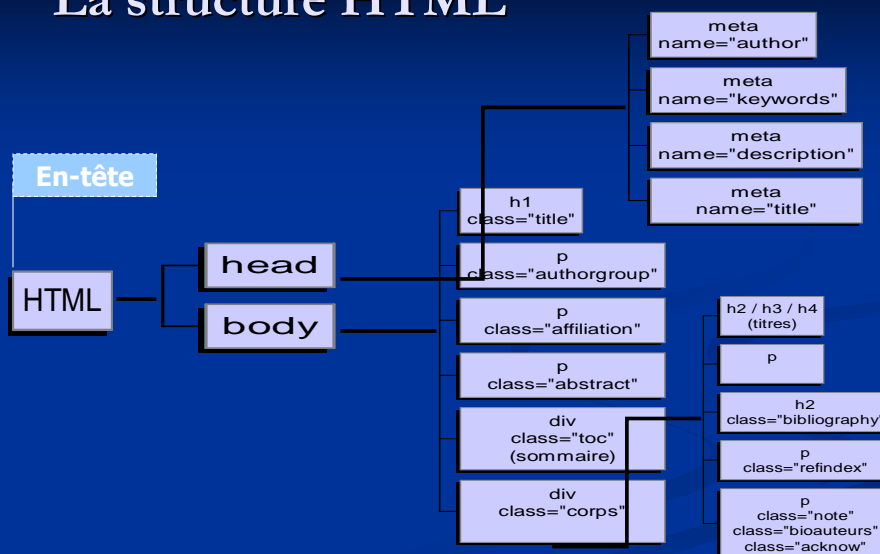
I ♥ VALIDATOR W3C XHTML 1.0

15/02/2011

C. Morel-Pair



# La structure HTML



15/02/2011

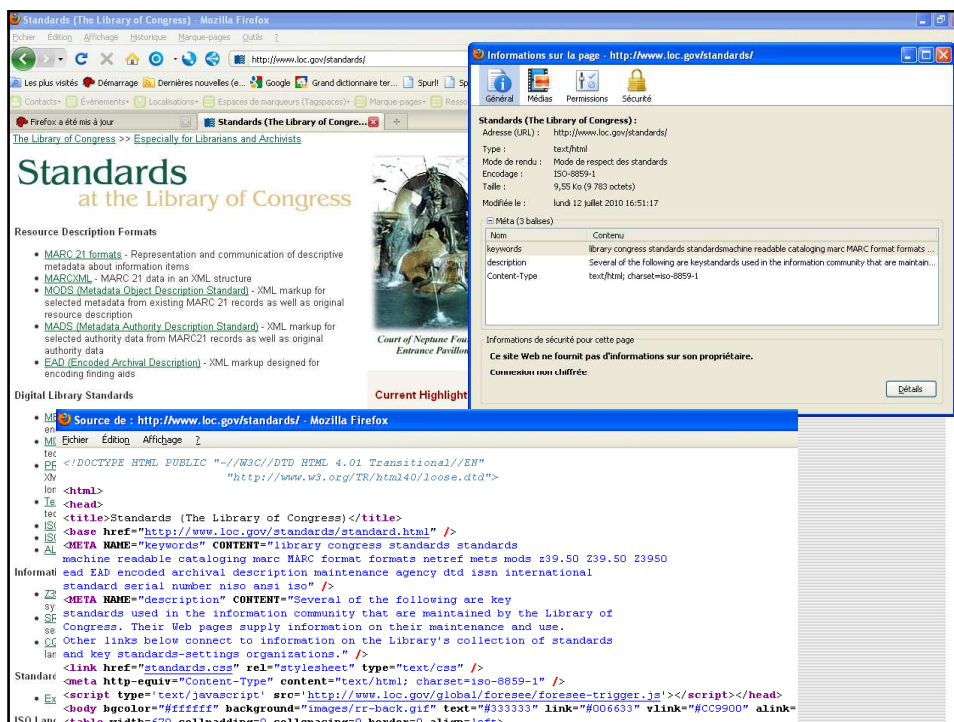
C. Morel-Pair

## 1 – Un premier standard de métadonnées Métadonnées HTML

- Dans la partie “head” de la page
- Descriptives du contenu
  - <title>
  - <meta name=“keywords” ou “description”
- Droits
  - <meta name=“author”, “publisher” ou “copyright”
  - + mécanismes satellites ajoutés : fichier robots.txt ...
- Techniques - relation
  - <meta content=“text/html” charset=“UTF-8”
  - <link rel=“stylesheet” ou “alternate”

15/02/2011

C. Morel-Pair



# Métadonnées (X)HTML Des usages ... des limites

## ■ Navigateurs

- Métadonnées techniques
  - Encodage de caractères
  - Feuilles de style
  - Fichiers alternatifs
- Métadonnées de droits

## ■ Annuaires (Yahoo, Cismef...) : indexation et catégorisation humaines à posteriori

## ■ Moteurs de recherche

- **title**
- **Description**, *keywords* ?
- **Pondéré** par autres critères
- Index enrichis avec métadonnées externes
- Le référencement passe aussi par la qualité des pages : structure et sémantique
  - "Référencement naturel"



# Google, recherche avancée

Google Advanced Search

Use the form below and your advanced search will appear here

Find web pages that have...

all these words:

this exact wording or phrase:

one or more of these words:  OR  OR

But don't show pages that have...

any of these unwanted words:

Need more tools?

Reading level:

Results per page:

Language:

File type:

Search within a site or domain:

Date:

Usage rights:

Where your keywords show up:

Region:

Numeric range:

SafeSearch:

Advanced Search

Page-specific tools:

Find pages similar to the page:

Find pages that link to the page:

©2011 Google

## Et les moteurs sémantiques, à facettes ... ?

Libra Paper Results - Mozilla Firefox

http://libra.msr.a.cn/paperresult.aspx?query=information-retrieval

Microsoft Libra Academic Search

Papers Authors Conferences Journals Communities

Search: information retrieval

Search Results: 1 - 10 of top 1000, totally 20030 (0 seconds)

**Related Authors:**

- Hector Garcia-molina
- Gerard Salton
- W. B. Croft
- Pattie Maes
- Ricardo Baeza-vates
- David Lewis
- Marti Hearst
- Ellen M. Voorhees
- Chris Buckley
- C. J. Van Rijsbergen

**Related Conferences:**

- SIGIR
- ISMIR
- CIKM

**Related Journals:**

- IPM

**Information Retrieval(1979)** (citation:1146)

C. J. Van Rijsbergen

...be concerned only with automatic **information retrieval** systems. automatic as opposed to manual and **information** as opposed to data or fact. unfortunately the word **information** can be very misleading. in the context of **information retrieval** (ir), **information**, in the technical meaning given...

**Modern Information Retrieval(1999)** (citation:1316)

Ricardo A. Baeza-vates Berthier A. Ribeiro-neto

...information retrieval (ir) has changed considerably in...

**Social information filtering: algorithms for automating ???word of mouth???(1995)** (citation:536)

Upendra Shardanand Pattie Maes

...use the systemand enter more **information**. four different algorithmsfor...

Conference: Computer Human Interaction - CHI

**A language modeling approach to information retrieval(1998)** (citation:387)



Ubiquitous Web  
domain

## 2 - XML

# Extensible mark-up language

```
1 <?xml version="1.0" encoding="utf-8"?>
2 <!DOCTYPE AmiereGrandParent SYSTEM "axe.dtd">
3
4 <AmiereGrandParent>
5   <nom>Eugene</nom>
6   <GrandParent>
7     <nom>Paul</nom>
8     <Parent>
9       <nom>Pierre</nom>
10      <File position="ainees">
11        <nom>Isabelle</nom>
12        <File>
13          <File>
14            <nom>Sylvain</nom>
15            <PetitFile>
16              <nom>Loic</nom>
17              <PetitFile>
18                <nom>Elodie</nom>
19                <PetitFile>
20                  <File position="cadetes">
21                    <nom>Julie</nom>
22                    <PetitFile>
23                      <nom>Antoine</nom>
24                      <PetitFile>
25                        <nom>Magalie</nom>
26                        <PetitFile>
27                          <File>
28                            <Parent>
29                              <GrandParent>
30                                <AmiereGrandParent>
```

- LE standard d'échange effectif aujourd'hui !
  - Maintenu par W3C
  - Soutenu par l'industrie (OASIS), et fondations
  - Intégré dans applications bureautique, toutes plateformes informatiques, logiciels, BD ...
- Web : XHTML, Web 2.0, Web services, Web sémantique ...
- Standards de publication
  - Tous domaines
  - Secteurs privés et publics
- Standards de métadonnées

15/02/2011

C. Morel-Pair

**COVER PAGES**  
Online resource for markup language technologies

**ABOUT**

### XML Applications and Initiatives

**Contents Listing for XML Applications and Industry Initiatives**

Many markup language applications in industry, government, and academia which began in the 1985-1990 timeframe as SGML applications have subsequently migrated to XML using XML DTDs, W3C XML Schemas, or other schema languages. Projects listed below are principally those which began as XML applications in 1990 or later. These applications and initiatives are not (here) categorized under fixed labels, but in many cases one will find related or similar topics grouped together. Should you fail to find a relevant TOC entry here, you may use the [website search facilities](#) or execute a Web browser search in the text of the [Site Index document](#) (load the document, then Edit -> Find or Search -> Find).

- W3C Specifications Documentation
- IEEE Standard DTD
- OASIS Specification DTDs and Tools
- Using XML for RFCs
- Text Encoding Initiative (TEI)
- Markup Language for Complex Documents (Bergen MLCD Project)
- Manuscript Access through Standards for Electronic Records (MASTER)
- XCES: Corpus Encoding Standard for XML
- Global Document Annotation Initiative (GDAI)
- Electronic Metadata for Endangered Languages Data (EMELD)
- Encoding and Markup for Texts of the Ancient Near East
- Electronic Text Corpus of Sumerian Literature (ETCSL)
- XML System for Textual and Archaeological Research (XSTAR)
- Perseus Project
- The Making of America II Project
- Metadata Encoding and Transmission Standard (METS)
- Harvard University E-Journal Archive Project
- Image Metadata Aggregation for Enhanced Searching (IMAGES)
- Encoded Archival Description (EAD)
- Encoded Archival Context Initiative (EAC)
- Linking and Exploring Authority Files (LEAF)
- STEP/EXPRESS and XML
- STEPml XML Specifications
- Atom Publishing Format and Protocol
- Channel Definition Format, CDF (Based on XML)
- RDF Rich Site Summary (RSS)
- Open Content Syndication (OCS)
- Web Modeling Language (WebML)
- Portable Site Information (PSI)
- XHTML and XML-Based HTML Modules

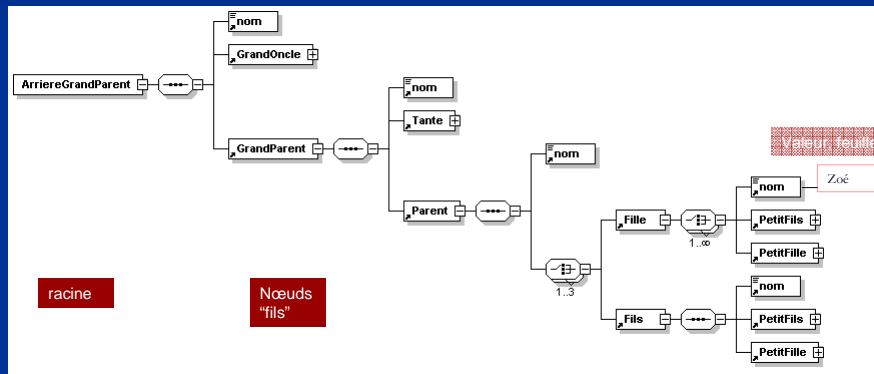
**HOSTED BY**  
**OASIS**

**SPONSORED BY**  
**IBM**  
**Microsoft**  
**ORACLE**  
**PRIMETON**  
普元·软件  
**Sun**  
microsystems

**XML DAILY NEWSLETTER**  
Receive daily news updates from Managing Editor, Robin Cover.  
[Subscribe >>](#)  
[View Archives >>](#)

# Le document XML

## Un arbre d'éléments



## Le document XML en bref

The screenshot shows the Altova XML Spy interface with an XML document open. The document structure is as follows:

```

<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE ArriereGrandParent SYSTEM "axe.dtd">
<ArriereGrandParent>
  <nom>Eugene</nom>
  <GrandParent>
    <nom>Paul</nom>
    <Parent>
      <nom>Pierre</nom>
      <Fille position="ainee">
        <nom>Isabelle</nom>
      </Fille>
      <Fils>
        <nom>Sylvain</nom>
        <PetitFils>
          <nom>Loic</nom>
        </PetitFils>
        <PetitFille>
          <nom>Elodie</nom>
        </PetitFille>
      </Fils>
      <Fille position="cadette">
        <nom>Julie</nom>
        <PetitFils>
          <nom>Antoine</nom>
        </PetitFils>
        <PetitFille>
          <nom>Magalie</nom>
        </PetitFille>
      </Fille>
    </Parent>
  </GrandParent>
</ArriereGrandParent>
  
```

Annotations in the image identify key parts of the XML document:

- En-tête** (Header): Points to the XML declaration and DOCTYPE declaration.
- racine** (root): Points to the root element, `<ArriereGrandParent>`.
- éléments** (elements): Points to the element nodes in the tree.
- fils** (sons): Points to the child elements of the root.
- valeur** (value): Points to the text content of an element, such as "Eugene".
- attribut** (attribute): Points to the `position="ainee"` attribute of the `<Fille>` element.
- base de données** (database): Points to the `SYSTEM "axe.dtd"` declaration.


The interface also shows a list of elements on the left, a list of attributes on the right, and a list of entities at the bottom right.

15/02/2006

Altova XML Spy v2006 sp1 U. Registered to Philippe FOURNIERET (ONIST - CHRS) ©1996-2005 Altova GmbH

Ln 22, Col 20 CAP NUM SCR

## XML, un choix efficient

- **Standard ouvert et libre, pérenne et interopérable**
    - Mode texte, Unicode et tous encodages
  - **“Métalangage” : base de différents modèles et applications**
    - Tous domaines, tous types de documents
  - **Efficace - comprendre**
    - Le document XML
    - Ses modèles (DTD, schémas) partagés
    - Des outils : modules de requête et traitement intégrés
    - **Un seul document → transformations et réutilisations multiples**
- 



15/02/2011

C. Morel-Pair

```

<?xml version="1.0" encoding="UTF-8" ?>
<ArriereGrandParent xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  <nom>Eugene</nom>
  <Coordonnees>
    <Adresse>21, rue Marguerite</Adresse>
    <CP>54000 Nancy</CP>
    <Telephone>0383505050</Telephone>
  </Coordonnees>
  <GrandOncle>
    <nom>Jean</nom>
  </GrandOncle>
  <GrandParent>
    <nom>Paul</nom>
    <Coordonnees>
      <Adresse>21, rue des Oeillots</Adresse>
      <CP>54000 Nancy</CP>
      <Telephone>0383505051</Telephone>
    </Coordonnees>
    <Tante>
      <nom>Irene</nom>
    </Tante>
  </GrandParent>
  <Parent>
    <nom>Pierre</nom>
    <Coordonnees>
      <Adresse>21, rue des Dalhias</Adresse>
      <CP>54500 Vandœuvre</CP>
      <Telephone>0383505052</Telephone>
    </Coordonnees>
    <Fille role="ainee">
      <nom>Isabelle</nom>
      <Coordonnees>
        <Adresse>21, rue Marguerite</Adresse>
        <CP>54000 Nancy</CP>
        <Telephone>0383505053</Telephone>
      </Coordonnees>
      <PetitFils>
        <nom>Loic</nom>
        <Coordonnees>
          <Adresse>21, rue de la Loire</Adresse>
          <CP>54000 Nancy</CP>
          <Telephone>0383505055</Telephone>
        </Coordonnees>
      </PetitFils>
      <PetiteFille>
        <nom>Elodie</nom>
      </PetiteFille>
    </Fille role="ainee">
  </Parent>
</ArriereGrandParent>

```

Eugene	21, rue Marguerite 54000 Nancy 0383505050
Jean	-
Paul	21, rue des Oeillots 54000 Nancy 0383505051
Irene	-
Pierre	21, rue des Dalhias 54500 Vandœuvre 0383505052
Isabelle	21, rue Marguerite 54000 Nancy 0383505053
Loic	21, rue de la Loire 54000 Nancy 0383505055
Elodie	-
Sylvain	21, rue des Lilas 54000 Nancy 0383505054
Julie	21, rue de Vic 54000 Nancy 0383505056
Antoine	-
Magalie	-

**Isabelle dit :**

Mes ancetres paternels sont :

- Eugene
- Paul
- Pierre

Mes descendants sont :

- Loic
- Elodie

**Julie dit :**

Mes ancetres paternels sont :

- Eugene
- Paul
- Pierre

Mes descendants sont :

- Antoine
- Magalie

**Transformation  
XSLT**

C. Morel-Pair

```
<dc:creator> BEN SALEM (Farah) </dc:creator>
```

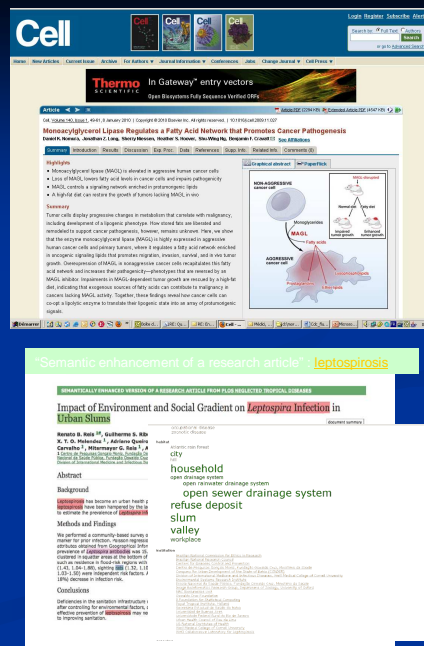
15/02/2011

[illegible]

## C. Morel-Pair



## C. Morel-Pair



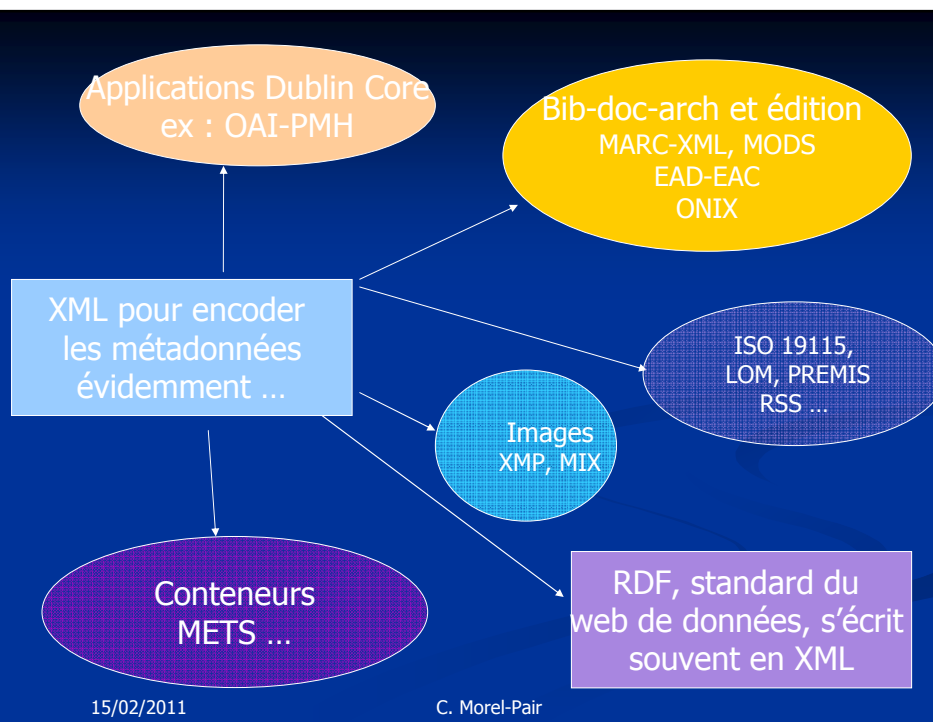
# Un modèle XML propriétaire : MSOffice 2007 et plus

- Expression par le monde économique de besoins d'interopérabilité et pérennité des documents produits
  - format MS "Office Open XML", standard propriétaire, normalisation longue
- Open Office : dès l'origine ...



15/02/2011

C





# Standards - Métadonnées

- *Métadonnées HTML*
- Jeu d'éléments **Dublin Core**
- Autres standards - sélection
  - MARC et XML
  - LOM ...



15/02/2011

C. Morel-Pair



## 1 – Standard Dublin Core

- Dublin Core Metadata Initiative (DCMI), 1995
  - Atelier à Dublin (Ohio) en 1995 (“OCLC/NCSA Metadata Workshop”)
  - Objectif : “Pallier les insuffisances des métadonnées HTML” pour donner une sémantique au Web [...] et améliorer la “resource discovery”
  - Par un standard généraliste
    - Une “description bibliographique simplifiée”
    - ... pour tout type de ressource, de producteur
  - Fondation active
  - <http://dublincore.org>

15/02/2011

C. Morel-Pair

# 1 - Dublin Core un vocabulaire en 4 modules

## ■ DC simple

- 15 éléments caractérisés par 10 propriétés
- Norme ISO 15836-2003

## ■ DC qualifié

- + “qualificatifs” (DCTERMS, “refinements, qualificatifs, extensions”) caractérisés

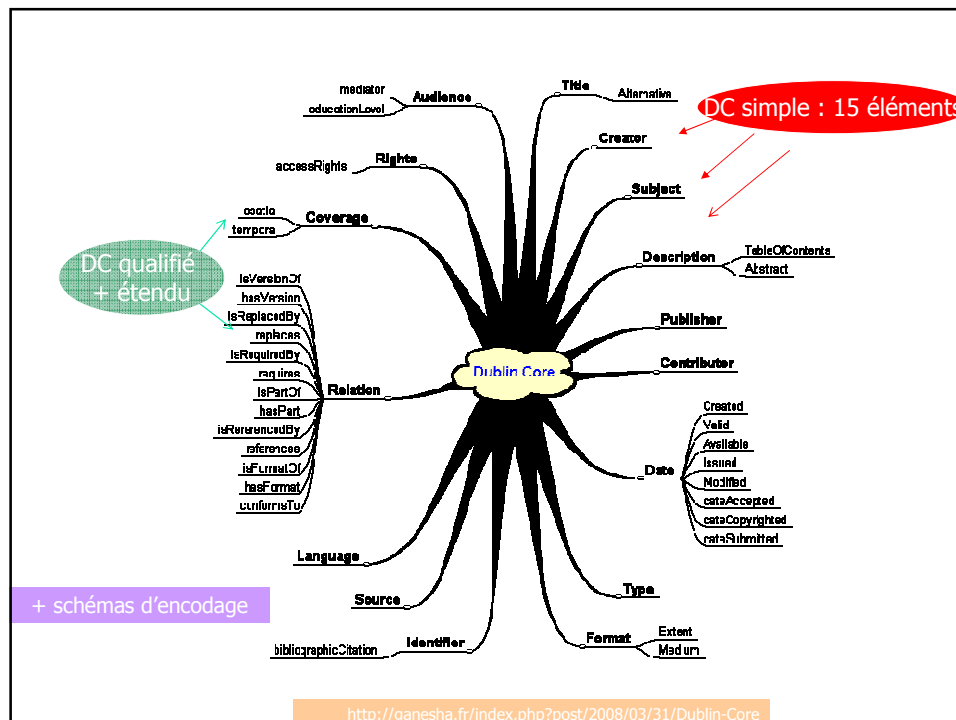
## ■ DC étendu

- + éléments issus d’autres standards (Learning Object Metadata, LOM)
- + éléments récents : gestion de collections ...

## ■ “Encoding schemes”, standardisation des valeurs


15/02/2011

C. Morel-Pair



<h1>Dublin Core simple</h1> <h2>Contenu intellectuel</h2>	<title> <subject> <description> <type> <relation> <coverage> <language>
<h2>Droits</h2>	<creator> <publisher> <contributor> <rights> <source>
<h2>Techniques</h2>	<date> <format> <identifier>

### Exemple (Dublin Core simple)



<b>Title</b>	Angelo Bilbao, April 24 1920
<b>Creator</b>	Sigler, Harold R.
<b>Description</b>	A portrait of Angelo Bilbao, a Basque resident of Boise, Idaho. The portrait was taken by R. Harold Sigler, who worked as a photographer in Boise for nearly 50 years. Among his photographs include a number of ethnic residents living in Boise, Idaho. For more information on the Sigler Collection at the Idaho State Historical Society click on the URL address located in the Collection field.
<b>Date</b>	1920
<b>Subject</b>	Basque Americans Boise (Idaho); Ada County Portraits Bilbao, Angelo
<b>Coverage</b>	United States--Idaho (State)--Ada--Boise
<b>Type</b>	Photographs
<b>Source</b>	Sigler Photograph Collection
<b>Identifier</b>	<a href="http://content.wsulib.wsu.edu/u?m16_9_382">http://content.wsulib.wsu.edu/u?m16_9_382</a>
<b>Format</b>	Original images were scanned as 300 dpi TIFF files on a Microtek 9600XL scanner. 72 dpi JPEG files were then added to the CONTENT database at the WSU Libraries.
<b>Rights</b>	To obtain a copy of this item, or request permission to publish, contact the Idaho State Historical Society Library and Archives, 2205 Old Penitentiary Road, Boise, ID. 837 12-8250, 208-334-3355.

Washington State University Libraries, digital collections

<http://content.wsulibs.wsu.edu/>

## DC, exemple éléments et qualificatifs

<b>Title</b>	Nom donné à la ressource, titre du document	<i>alternative</i> : autre titre (traduit, abrégé ...)
<b>Subject</b>	Mots-clés, codes de classement, sujet. Utilisation de vocabulaires ou classifications contrôlés recommandée (LCC, UDC,DDC LCSH, Mesh, TGN)	
<b>Description</b>	Description du document: résumé, table des matières, texte libre	<i>abstract</i> <i>tableOfContents</i>

15/02/2011

C. Morel-Pair

## DC exemple élément “relation” et qualificatifs

<b>Relation</b>	Lien vers une ressource liée. Il est recommandé d'utiliser une dénomination formelle des ressources (URI)
<i>Qualificatifs</i>	<i>isFormatOf / hasFormat</i> <i>isVersionOf / hasVersion</i> <i>isReplacedBy / replaces</i> <i>isPartOf / hasPart</i> <i>isReferencedBy / references</i> <i>isRequiredBy / requires</i> <i>conformsTo</i>

15/02/2011

C. Morel-Pair

## Les “schémas d’encodage” renseignent les éléments de manière homogène et non-ambigüe

- Valeurs issues de **ressources reconnues**
  - Thésaurus, classements et ontologies...
    - Ex DCMI : LCSH, MESH, LCC, DDC, TGN, DCMItypé
  - Autorité personnes ou organismes
  - Codes ISO, RFC ...
    - Ex: codes langues et pays
- **Modèles de valeurs** attendues
  - Ex : structure d’une date, de coordonnées géographiques
  - **Identifiants uniformes** des ressources : **URI** (DOI, ISSN ...)

15/02/2011

C. Morel-Pair

## Dublin Core, intérêt et limites

- Consensus international et interprofessionnel
  - Sémantique “commune” interdisciplinaire ; “multilinguisme”
  - Simplicité – 2 niveaux d’usage
  - Extensibilité et flexibilité
  - Normalisation ET évolutivité
- **largement adopté aujourd’hui**
- Description généraliste, pour la recherche d’informations
    - Métadonnées techniques et de structure limitées
  - Ressource plutôt électronique et isolée
  - Architecture peu contraignante
- ... **en association avec d’autres jeux**  
“**profils d’application**”

15/02/2011

C. Morel-Pair

## Détour

# “URI”, identifiants uniformes des ressources

- Une “chaîne de caractères”
  - correspondant de manière non ambiguë à une ressource
  - dans un contexte spécifié “espace de nommage”
- permet l'identification de la ressource
- et éventuellement
  - Sa localisation ...
  - L'accès
    - // droits ! ...
- “uniforme” = syntaxe particulière

Objet, numérique ici

issn: 1951-6827

doi: 10.86/1292

Possible :

1 ress. = plusieurs ident  
1 ident = x fichiers

http: irevues.inist.fr/index.php?id=68

15/02/2011

C. Morel-Pair

## Principaux URI pour les ressources numériques

ISSN, ISBN, ISRN ... électroniques

URL

Mais un jour ...

The requested URL could not be retrieved

PURL

<http://purl.org/dc/elements/1.1/>

Handle

ex : [hdl:2042/344](http://hdl.handle.net/2042/344)

DOI

doi:10.1045/jun  
e2005-jantz

Système ARK

Résolveurs de liens

<http://hdl.handle.net>  
<http://www.euroref.org/>  
<http://gallica.bnf.fr/ark:>

“métas-URI”  
projets URN et info-uri



# URI : interopérabilité débutante avec le web

The screenshot shows a Mozilla Firefox browser window. The address bar displays the URI: `hdl:2042/344`. The page content includes a search bar, a title 'Au XXe siècle, le temps des femmes et les saisons de leur corps', and a list of Dublin Core metadata fields such as 'URI', 'Titre', 'Auteur', 'Mots', 'Image', 'Société', 'Séjour', 'Corps des femmes', 'Breast', 'Society', 'Woman's', 'Image', 'Woman's body', 'Éditeur', and 'Date'. A text box on the left side of the slide reads: 'Firefox, add-on pour la résolution "directe" par le navigateur'.

15/02/2011

## Où trouve-t-on du Dublin Core ? formats d'encodage

- Formats de présentation : bureautique et pdf, tableaux HTML ...
- Plus interopérable ...
  - **HTML, XHTML** (partie head, élément <meta ...>
    - Pas (encore?) analysé par les grands moteurs ...
    - Mais utilisé par certaines applications : annuaires, Zotero ...
  - **XML** : métadonnées DC dans un autre fichier - "notice"
    - Des usages importants pour mutualiser : bibliothèques numériques, notice minimale du protocole OAI-PMH, métadonnées TEF, portails de l'information publique ...
  - **En RDF, Dublin Core est une ontologie reconnue du web sémantique**, dans une ressource (HTML) ou à côté

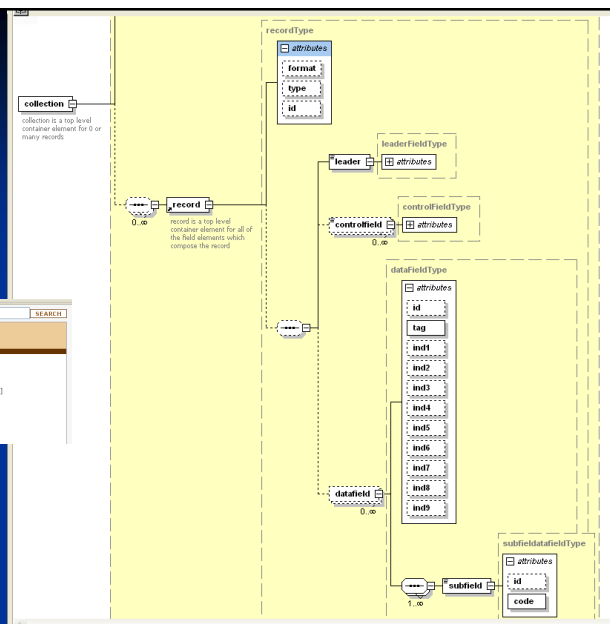
15/02/2011

C. Morel-Pair



## 2 – standards bibliographiques

Habiller \*MARC en XML :  
**MarcXchange**  
2008



**MARCXML**

MARC 21 XML Schema  
Official Web Site

Pionnier, 1999



## MarcXchange Une notice UNIMARC (ABES)

```
<controlfield tag="001">RU/NLR/BIBL/128053</controlfield>
<controlfield tag="005">20050608165836.0</controlfield>
```

élément contenant les zones  
001 à 009 et 00A à 00Z

```
<datafield tag="105" ind1="1" ind2="1">
  <subfield code="a">y |||||</subfield>
</datafield>
```

élément contenant les  
zones de données

```
<datafield tag="200" ind1="1" ind2="1">
  <subfield code="a">The Liberal Party in rural England, 1885-1910</subfield>
  <subfield code="e">radicalism and community</subfield>
  <subfield code="f">Patricia Lynch</subfield>
</datafield>
```

élément contenant les  
sous-zones

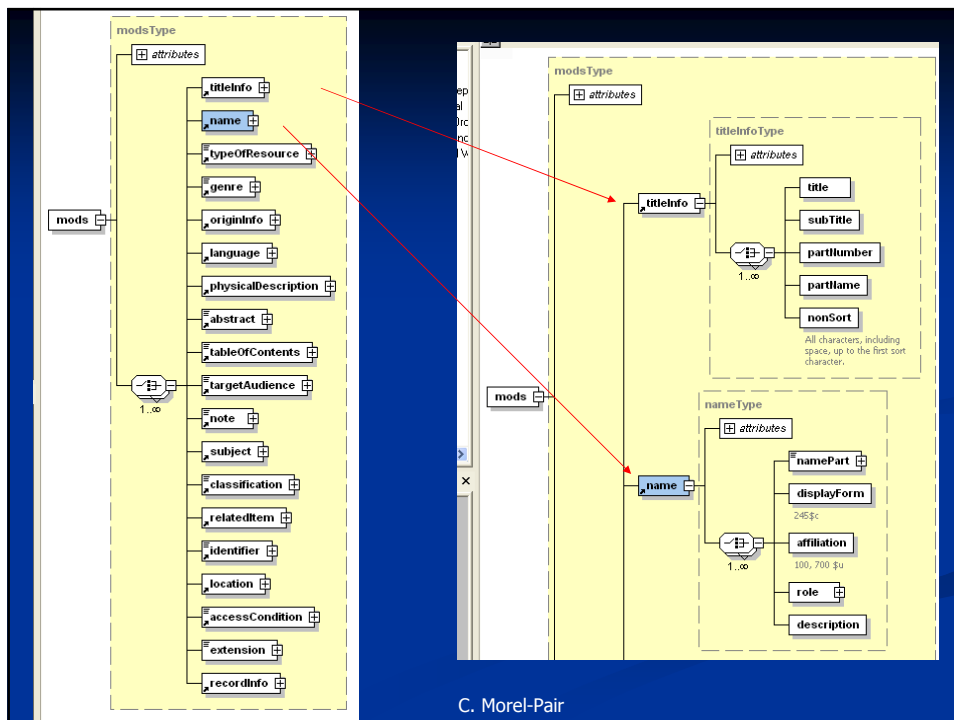
Format pivot imports-export  
web domain

Source : [http://www.abes.fr/abes/documents/reseau/journees\\_reseau/02c\\_ISO2709\\_MarcXchangeV2.ppt](http://www.abes.fr/abes/documents/reseau/journees_reseau/02c_ISO2709_MarcXchangeV2.ppt)

- Sous-ensemble d'éléments MARC verbalisés, "suffisant pour créer un enregistrement complet"
  - + convivial, + simple et + orienté document électronique que MARC-XML
  - + riche que Dublin Core, + proche habitudes bibliothèques que ONIX
  - flexible, bon niveau de granularité
  - Né et stabilisé en 2003 – Bibliothèque du Congrès
- Applications variées : collections textes, images, AV, partitions, annuaires de sites Web, bâtiments ...

15/02/2011

C. Morel-Pair



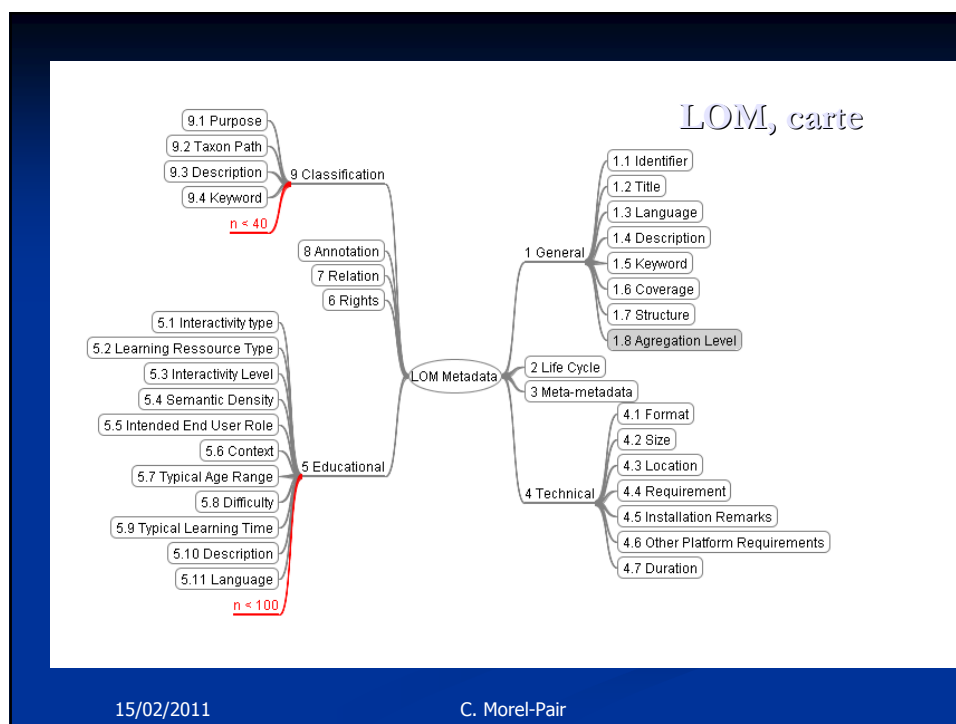
C. Morel-Pair

# MODS, pivot de DAPHNE



## 4 - LOM, Learning Object Metadata

- Développement de la FOAD et du nombre des ressources pédagogiques numériques disponibles
  - Accessibilité, évaluation, pérennité et interopérabilité : pour de nouveaux usages, dans le respect des droits
- LOM, IEEE, 2002 - standard international examiné par l'ISO
  - Profil d'application LOMFR, norme AFNOR décembre 2006
    - LOMFR-sup (2007) et LOMFR-scol (2010)
- Utilisateurs
  - Editeurs commerciaux
  - Communautés enseignement formation
    - Universités Virtuelles, ORI-OAI, Correlyce ...
    - Outils : EducaMeta, LOMPad, LOMfr-Pad, ORI-OAI ...



### Signets

Signets sélectionnés par les bibliothèques universitaires

Rechercher un signet

Flux rss tous les signets

SCIENTES ET TECHNIQUES

Sciences médicales (581)

Signets 1 à 10 / 581 | Page 1 / 59 | Aller à la page

Flux rss de cette thématique

Centre Français d'Evidence Based Dentistry (CFEBD)

Basé à l'UPR d'odontologie de Toulouse, le CFEBD a pour objectif de fournir aux chirurgiens-dentistes, aux chercheurs et aux patients des informations récentes sur les données avérées de la science. Le CFEBD travaille en collaboration avec le groupe Cochrane pour la santé bucco-dentaire et traduit notamment les résumés des revues systématiques Cochrane.

Classification DEWEY : 617.6

Langues : Français

Pays : France (métropolitaine)

Mots-clés : Odontologie, Stomatologie, EBM

Sélectionné par : Université Paul Sabatier - Toulouse 3

Propriété intellectuelle : UPR d'odontologie, Université Paul Sabatier (Toulouse 3)

Éditeur : L'Éditeur

Ajouter un commentaire

Agence régionale de santé

http://www.ars.midi-pyrenees.sante.fr/Internet/midi-pyrenees\_0...

Ouverte début avril 2010, l'Agence régionale de santé (ARS) de Midi-Pyrénées met en œuvre la politique régionale de santé, en coordination avec les partenaires et en tenant compte des spécificités de la région et de ses territoires.

Classification DEWEY : 614

Langues : Français

Pays : France (métropolitaine)

Mots-clés : politique sanitaire, santé publique

Sélectionné par : Université Paul Sabatier - Toulouse 3

Éditeur : L'Éditeur

Ajouter un commentaire

Cancéropôle du grand Sud-ouest

http://www.canceropole-gso.org/

Le cancéropôle contribue à coordonner la recherche en cancérologie au plus haut niveau d'excellence internationale, en particulier dans les domaines nouveaux issus de la révolution génomique et dans ceux des sciences sociales et économiques.

Classification DEWEY : 610 - 615.5 - 616.994

Langues : Français

Pays : France (métropolitaine)

Mots-clés : cancer, oncologie

Sélectionné par : Université Paul Sabatier - Toulouse 3

Éditeur : L'Éditeur

Ajouter un commentaire

Signets

Actualités

Participants

Présentation

Entrepôt OAI

Espace contributeurs

CERIMES

abes

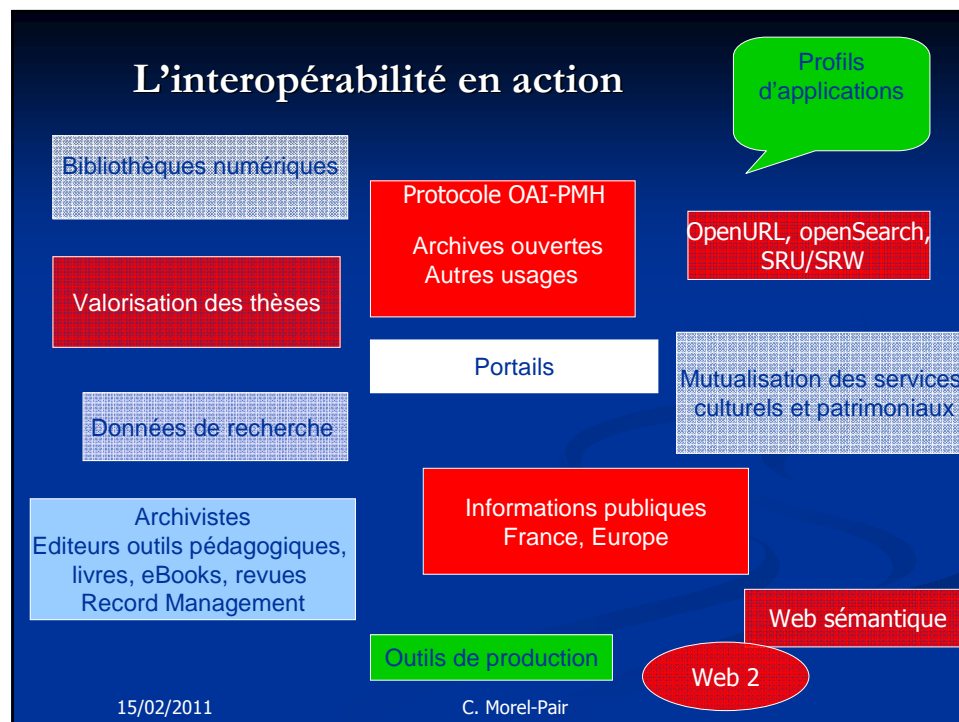
Agence Interuniversitaire de Management Hospitalier

Université Paul Sabatier

- Métadonnées ? Généralités et standards
- L'interopérabilité en action - exemples
  - Métadonnées et protocole OAI-PMH
    - Pour les archives ouvertes scientifiques et d'autres usages
  - Métadonnées dans un projet de valorisation des thèses : TEF
  - Mutualisation des ressources enseignement supérieur : ORI-OAI ... et ses métadonnées

15/02/2011

C. Morel-Pair



15/02/2011

C. Morel-Pair



# 1 – Archives ouvertes et protocole OAI-PMH

- Un contexte stratégique, le mouvement du libre accès aux publications scientifiques – Open Access **“OA”**
  - **Accès, visibilité, usages**
- Conférences de Budapest, Bethesda, Berlin
  - Adhésion de nombreux pays et organismes de recherche
- **Voie d’or : revues avec peer-review en libre accès**
  - DOAJ : ~6000 revues aujourd’hui, 500 000 articles
  - 783 avec facteur d’impact JCR 2009
- **Voie verte : auto-soumission dans des archives ouvertes** – Open Archive Initiative **“OAI”**
  - ~ 1500 archives ouvertes aujourd’hui, thématiques ou institutionnelles – types de documents variés



15/02/2011

C. Morel-Pair



## Protocole OAI-PMH

- **Quelle interopérabilité** pour ces “archives ouvertes” ?
  - “Web invisible” (BDs) et “architecture distribuée”
  - **Protocole OAI-PMH**, 2001, Convention de Santa Fe
    - Protocole de “**moissonnage**” / collecte des métadonnées des “**entrepôts**” des archives ouvertes - Échanges machines
- ... **Protocole simple et ouvert** → **autres usages**
  - Revues en ligne
  - Portails patrimoine-culture et administration
  - Mutualisation de “signalements” dans divers projets

15/02/2011

C. Morel-Pair



## Archives ouvertes, exemples

“Vues” du réservoir principal : TEL, HAL-SHS, HAL-INRIA ...

Interopérabilité avec ArXiv, Pubmed

## Université de Liège et accès libre

- Source : [présentation](#) de Paul Thirion, Metz, 2010-09
- “Serial price crisis”
  - Concentration des éditeurs - coûts des abonnements
    - Acquisitions ULg > 2 millions euros / an
    - Utilisateur – payeur différents ...
  - Délais de publication et tyrannie du facteur d’impact
- Alors que en bibliothèque universitaire
  - Compétences documentaires et technologiques
  - Rôle de passeurs de savoir = “construire”, aussi
  - Philosophie “open”, partagée avec direction ULg

15/02/2011

C. Morel-Pair



# Université de Liège et libre accès Voie d'or et voie verte



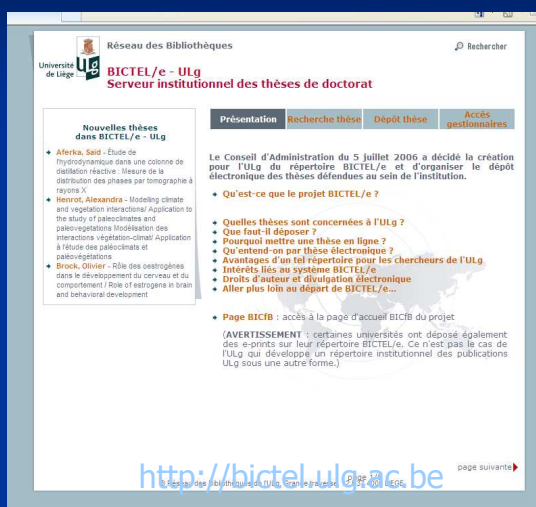
- **PoPuPs**,  
Portail de  
Publication des  
Périodiques  
scientifiques de  
l'ULg, (2005)  
[http://popups.  
ulg.ac.be](http://popups.ulg.ac.be)

- ~20 revues,  
4000 articles
- Outil Lodel

15/02/2011

C. Morel-Pair

## ULg - BICTEL/e



- Dépôt obligatoire  
des thèses de  
doctorat sous forme  
électronique (2006)
- Alimente  
BICTEL/e national
- Interopérable avec  
autres projets thèses
  - ND LTD
  - DART-Europe

<http://bictel.ulg.ac.be>

15/02/2011

C. Morel-Pair



### C. Morel-Pair

### C. Morel-Pair

# ORBi : un succès

- Nombre de dépôts et de consultations-téléchargements
- Au-delà de l'obligation, de la gestion de projet :
- Outil de soumission orienté utilisateur ++
- Services à valeur ajoutée
  - Pérennisation des documents
  - Ajout automatique statistiques et bibliométrie → métadonnées
  - Bon référencement : outils libre accès, moteurs du web et catalogues commerciaux
  - Edition de listes de publications personnalisables
  - Ouverture sur le web : ressources externes, widgets → interopérabilité
  - Liens actifs avec l'annuaire ULg
- Outil : DSpace très "customisé", voir "[fonctionnalités et services](#)"

15/02/2011

C. Morel-Pair

**Jorum Learning to Share**

Home | Get Started | Community | Support | News | About Us | Deposit

You are here: Jorum Search Home > Search Results

**Calculating Medicines for Children**

Show full item record

Share | | | |

**Title:** Calculating Medicines for Children  
**Author:** Gill Langmack  
**Description:** This RLO is aimed at providing examples of how to calculate medicines, specifically for children but it relates well to any aspect of calculating medicines.  
**Keywords:** drugs; pharmacy; drug calculations; paediatrics; children; clinical skills; numeracy; pharmacology  
**Persistent Link:** <http://open.jorum.ac.uk/xmlui/handle/123456789/396>  
**Date:** 2010-01-13

**Package Download** **Package Preview**

**CC Licence:** This item is licensed under a [Creative Commons](#) licence. By clicking on any of the above buttons(s) you agree to follow the terms of that licence.

**Files in this item**

Files	Size	Format	View
childcalc_notes/ovsync.xml	158bytes	Unknown	<a href="#">View</a>   <a href="#">Download</a>
childcalc2.html	7.981Kb	HTML	<a href="#">View</a>   <a href="#">Download</a>
childcalc3.html	7.989Kb	HTML	<a href="#">View</a>   <a href="#">Download</a>

Outils : intraLibrary for JorumUK and a modified version of DSpace for JorumOpen.



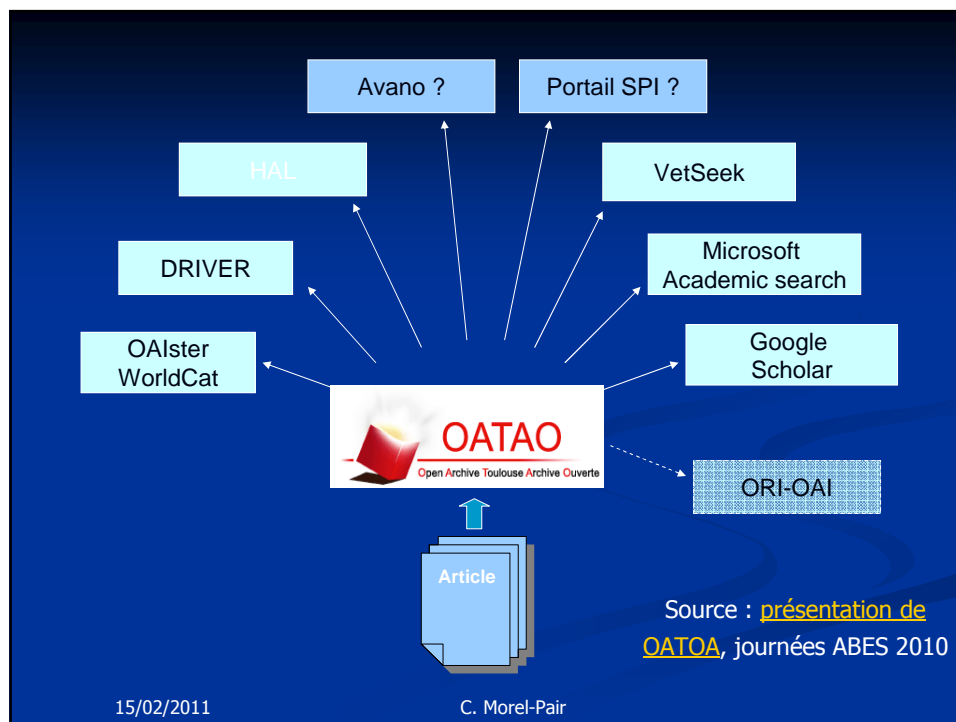
## OATAO

### Pourquoi un projet local ?

- Archive ouverte institutionnelle multi-établissements, 2008-01 Toulouse
- **Visibilité, qualité et souplesse**
  - LRU : pilotage et politique de valorisation de la recherche → vitrine
  - Partir des besoins et pratiques des chercheurs et mettre les compétences documentaires au cœur du processus
  - Evolutivité – incitation locale au dépôt
  - Lier les publications et d'autres ressources : brevets, CV, thèses ...
  - Possibilités de collaborations variées, choisies et évolutives
- **Valeur ajoutée**
  - Intégration des statistiques aux métadonnées (téléchargement ... citation)
  - Export de listes de publications ... incitation au dépôt
- **Interopérabilité** : HAL, moissonnage par d'autres "entrepôts"

15/02/2011

C. Morel-Pair



15/02/2011

C. Morel-Pair

# Interopérabilité des archives ouvertes

## Protocole OAI-PMH

- Les moissonneurs OAI-PMH sont des moteurs spécialisés collectant les métadonnées d'un ensemble d'entrepôts OAI pour permettre une interrogation fédérée
  - Avec ou non services ajoutés
- Moissonneurs généralistes
  - [OAIster](#)
  - [Citebase Search](#) et étude de l'impact
  - [Scientific commons](#)
- Moissonneurs spécialisés domaines scientifiques
  - Avano, Perseus, ISIDORE ...

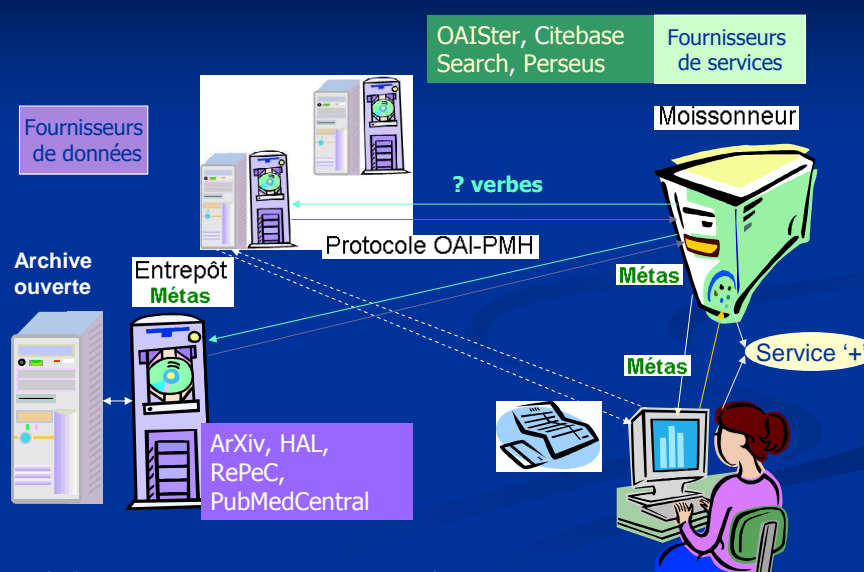


[citebase Search](#)

15/02/2011

C. Morel-Pair

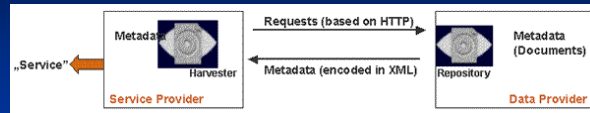
## Protocole OAI-PMH



15/02/2011

C. Morel-Pair

# OAI-PMH, les verbes de requête



## Les commandes

(ex avec le dépôt OAI de revues.org : <http://oai.revues.org>)

- Identifier le dépôt : <http://oai.revues.org/?verb=Identify>
- Lister les enregistrements (les articles, dans notre cas) : [http://oai.revues.org/?verb=ListRecords&metadataPrefix=oai\\_dc](http://oai.revues.org/?verb=ListRecords&metadataPrefix=oai_dc)
- Lister les « sets » (les revues, dans notre cas) : <http://oai.revues.org/?verb=ListSets>
- Lister les formats disponibles : <http://oai.revues.org/?verb=ListMetadataFormats>
- Lister les identifiants : [http://oai.revues.org/?verb=ListIdentifiers&metadataPrefix=oai\\_dc](http://oai.revues.org/?verb=ListIdentifiers&metadataPrefix=oai_dc)

19/02/2011 C. Morel-Pair

Identify

ListSets

ListFormats

ListIdentifiers

ListRecords

GetRecord

+ paramètres  
format métadonnées  
date  
nb d'enregistrements  
Identifiant enregistrement

Un exemple dans  
[I-Revues](#)

## Moissonneurs : résultat de requête OAIster "ancien"

Home Search Browse Institutions Help

Your search was in the **Title** field for "metadata".

You found **818 records**.

- \* [Revise your search](#) to retrieve fewer records.
- \* View your results, starting with **records 1 to 10 of 818**.

Results by Institution

Sort by: date descending

Jump to Records: 1 | 11 | 21 ... 81 ... 161 ... 241 ... 321 ... 401 ... 481 ... 571 ... 651 ... 731 ... 811

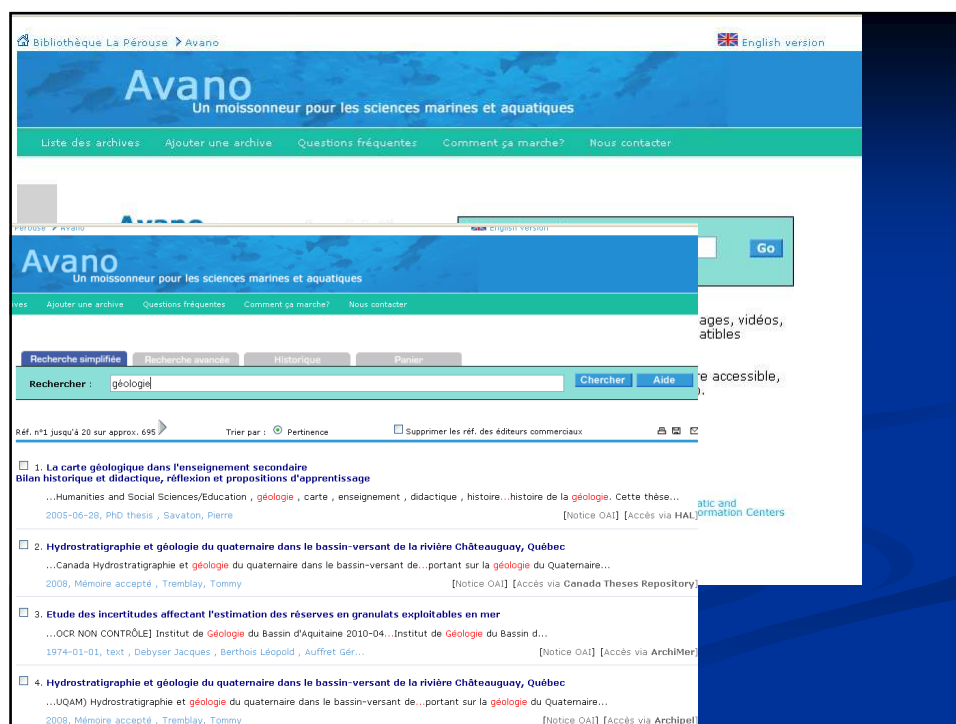
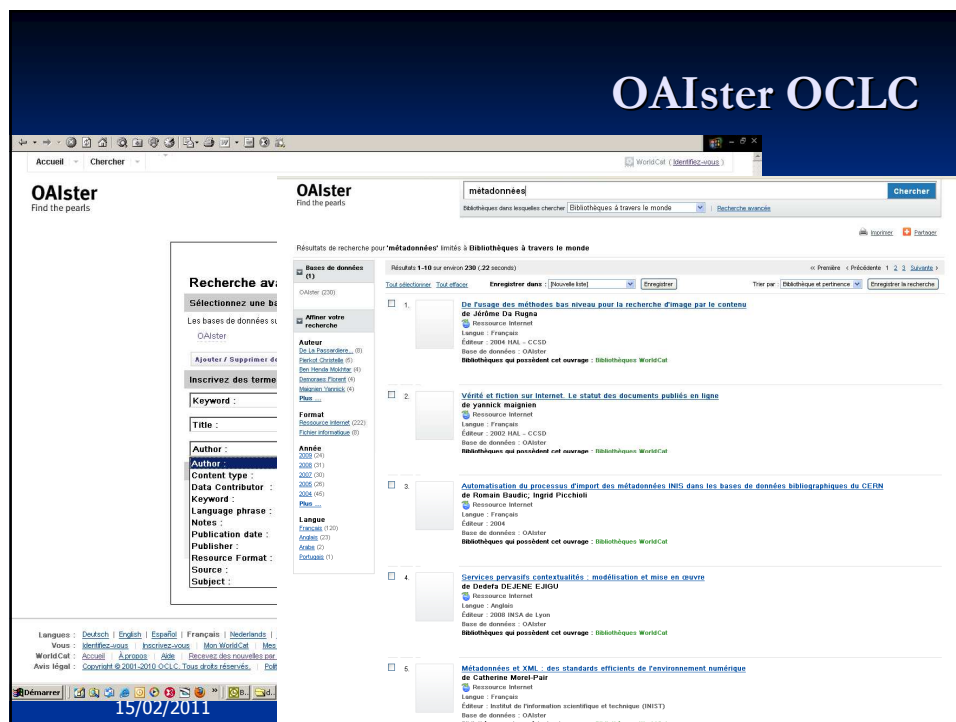
Next 10 Records

**Record 1 of 818**

Title	Digital collection management software employed by libraries and museums: examination of <b>metadata</b> semantic mapping functionality
Contributor	Park, Jung-ran
Contributor	Park, Sang Joon
Year	2005-09-06T20:37:49Z
Resource Type	Presentation
Resource Format	91252 bytes,application/pdf
Language	en_US
Note	A growing number of organizations are building digital collections using both commercial digital collection management software such as CONTENTdm, Encompass, etc., and open source software such as Greenstone and D-Space. This rapidly growing number of distributed digital collections has brought to the fore the critical issues of resource discovery and sharing across these collections. The goal of this project is to examine the functionality of <b>metadata</b> creation and mapping and the configuration of digital collection management software. This goal relates to the issue of semantic interoperability of concept representation across digital collections. For this, this project aims at examining how digital collection management software provides a mechanism for semantic mapping either between different <b>metadata</b> schemes such as Dublin Core (DC) and MARC or between cataloger-defined field names and a given <b>metadata</b> scheme such as DC. As a first step, we will examine features related to <b>metadata</b> semantic mapping of CONTENTdm software, which provides a feature that allows for catalogers to map cataloger-defined field names onto DC <b>metadata</b> elements.
Note	IST Research Day 2005 posters;no. 105F
Subject	<b>Metadata</b> ; Data conversion; Electronic libraries
URL	<a href="http://hdl.handle.net/1860/524">http://hdl.handle.net/1860/524</a>
Institution	DSpace at Drexel, Electronic Theses and Dissertations


**Record 2 of 818**

Title	Developing a Canadian <b>Metadata</b> Profile for Institutional Repositories
Contributor	Jordan, Mark





☐ 4. **Hydrostratigraphie et géologie du quaternaire dans le bassin-versant de la rivière Châteauguay, Québec**

 Accéder à la ressource

Dates: 2008 , 2010-12-10T14:25:21Z

Auteurs: Tremblay, Tommy

Descriptions: Une étude portant sur la **géologie** du Quaternaire dans la portion canadienne du bassin-versant de la rivière Châteauguay a été entreprise dans le cadre d'un programme d'étude sur les eaux souterraines mené conjointement par la Commission géologique du Canada, l'INRS-ETE (Institut national de recherche scientifique-Eau, Terre et Environnement) et le Ministère de l'environnement. Les objectifs principaux ont porté sur les relations entre la cartographie des dépôts meubles et la recharge à l'aquifère fracturé, ainsi que sur l'exploitation des données de sous-surface pour la conception d'une carte en 3 dimensions de la **géologie** quaternaire (hydrostratigraphie). Une méthode novatrice par grille (grid-based), la méthode de calcul relatif, a été développée pour la cartographie en 3 dimensions. Des travaux de terrain ont été réalisés, dont des forages dans les argiles à la mini-foreuse, de la cartographie de surface et des études sédimentologiques (particulièrement dans les eskers). Parallèlement, des études géologiques portant sur l'histoire de la glaciation et de la déglaciation ont été entreprises. La géomorphologie glaciaire, les marques d'érosion glaciaire ont été étudiées pour compléter les connaissances sur l'histoire des écoulements glaciaires dans la région. Les stries indiquent un mouvement glaciaire principal dans la région; pu localement. Un site paléontologique important (li retrouvé des fossiles de plantes arctiques-alpines). Des datations au carbone 14 ont été effectuées, important (environ 800 ans pour Hiatella arctica). Des datations par IRSL (Infrared-Stimulated Luminescence) post-datant la déglaciation. Enfin, les Champs de la Rivière ont suggéré de nouveaux détails sur la déglaciation.

L'AUTEUR : **Géologie**, Quaternaire, Châteauguay, Champlain.

Identifiants: oai:www.archipel.uqam.ca:1814  
<http://www.archipel.uqam.ca/1814/1/M10248.pdf>  
<http://www.archipel.uqam.ca/1814/2/m10248.JPG>

Type: Mémoire accepté , PeerReviewed

Formats: application/pdf , other

Relations: <http://www.archipel.uqam.ca/1814/>

Archive: Archipel: Archive de Publications Electroniques, U

Avano, un moissonneur pour les sciences marines et aquatiques - Mo...

Fichier Édition Affichage Historique Marque-pages Outils ?

<http://www.ifremer.fr/av...>

Contacts\* Événements\* Localisations\* Espaces de marqueurs (Tagspaces)

Avano, un moissonneur pour les scie...

Please select an URL:

☒ <http://www.archipel.uqam.ca/1814/1/M10248.pdf>

☒ <http://www.archipel.uqam.ca/1814/2/m10248.JPG>

## OAI-PMH, les entrepôts

### ■ Contenu minimal d'un entrepôt de métadonnées

- Dublin Core – XML
- Dans une base de données (\*SQL, ODBC ...)

### ■ Autres formats moissonnables

- MARC-XML, MODS, ETD-MS
- METS, EAD, DIDL ...
- ... Tout format XML

### ■ Outils de création

- Plateformes de publication numérique: EPrints, DSpace, CDSware, FEDORA, HAL ... ContentDM, Divvalib ...
- Logiciels documentaires et archivistes : de plus en plus

http://documents.hdl\_2042\_28804 La Météorologie - N°66 - Août Réduire

```

- <record>
- <header>
  <identifier>oai:documents.irevues.inist.fr:2042/28830</identifier>
  <datestamp>2009-11-20T06:28:18Z</datestamp>
  <setSpec>hdl_2042_28804</setSpec>
</header>
- <metadata>
  <oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/oai/2.0/oai_dc.xsd">
    <dc:creator>BOE, Julien</dc:creator>
    <dc:date>2009-10-13T12:15:49Z</dc:date>
    <dc:date>2009-10-13T12:15:49Z</dc:date>
    <dc:date>2009</dc:date>
    <dc:identifier>http://hdl.handle.net/2042/28830</dc:identifier>
    <dc:description>
      La complexité des mécanismes liés au cycle de l'eau continental joue un rôle important dans la disparité actuelle des résultats des modèles climatiques sur la France et donc dans les incertitudes quant aux impacts du changement climatique d'origine anthropique. Cet article présente les rétroactions qui sont en jeu et souligne l'importance pour les modèles de climat de représenter correctement le cycle hydrologique du climat présent. Cela est particulièrement critique pour prévoir l'effet du changement climatique en été sur la France et plus largement sur une large zone intermédiaire entre l'Europe du Nord et l'Europe du Sud. Pour ces régions, une meilleure compréhension des changements de circulation atmosphérique sur l'Atlantique nord est conjointement nécessaire. Complété par une étude d'impact, l'article souligne que, en France, dans l'état actuel des connaissances, des impacts importants sont à attendre sur la ressource en eau.
    </dc:description>
    <dc:description>
      Discrepancies between climate models results over France are largely due to the complexity of mechanisms involved in the continental water cycle and thus on the ensuing uncertainties in the impact of the anthropological climate changes. This article presents the feedback mechanisms involved and emphasizes the importance for climate models to properly represent the hydrological cycle of today's climate. This is particularly critical to forecast the effect of climate change in summer over France, and more broadly, over the wide transition zone between Northern and Southern Europe. For these areas, a better understanding of the changes in the atmospheric circulation over the North Atlantic is also necessary. Ending with an impact study, this paper highlights that with today's understanding, important consequences are to be expected on water supply over France.
    </dc:description>
    <dc:format>1678447 bytes</dc:format>
    <dc:format>application/pdf</dc:format>
    <dc:language>fr</dc:language>
    <dc:publisher>Société météorologique de France, Paris (FRA)</dc:publisher>
    <dc:rights>
      Document réservé aux abonnés - Licence d'utilisation : http://documents.irevues.inist.fr/utilisation.jsp
    </dc:rights>
    <dc:source>
      La Météorologie [ISSN 0026-1181], 2009, Série 8, N° 66 ; p. 38-48
    </dc:source>
    <dc:subject>climatologie</dc:subject>
    <dc:subject>variation du climat</dc:subject>
    <dc:subject>cycle hydrologique</dc:subject>
    <dc:subject>influence des activités humaines</dc:subject>
    <dc:subject>changement climatique</dc:subject>
  </oai_dc:dc>

```

Header : identifiant, date création, set

Metadata, ici dublin Core

'creator'

'date(s)'

'identifier'

'description(s)'

'format'

'language'

'publisher'


'rights'

'source'

'subject'

## Une ressource peut avoir plusieurs représentations

[http://memory.loc.gov/cgi-bin/oai2\\_0?verb=ListMetadataFormats](http://memory.loc.gov/cgi-bin/oai2_0?verb=ListMetadataFormats)





[Afro-American music sheet](#)  
(collection de ressources ici)

oai:lcoa1.loc.gov:lccn/0528551

- Dublin Core
- OAI MARC
- MODS
- MARC 21 Slim

Standards adaptés à contenus et objectifs

- ORBI (4 dont DIDL)
- Bictel-e ULG (4 dont ETD-MS) →

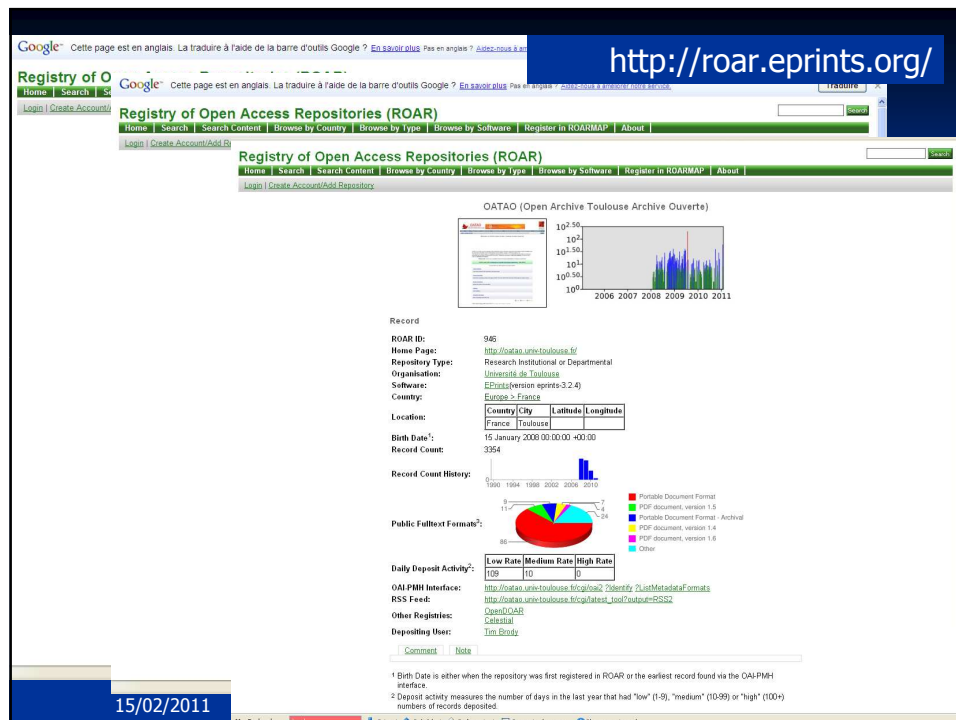
15/02/2011 C. Morel-Pair

# Visibilité des entrepôts OAI-PMH

- Outre le moissonnage ...!
- Des répertoires très informatifs, pas toujours exhaustifs - exemples
  - <http://www.opendoar.org/index.html>
  - <http://roar.eprints.org/>
  - <http://oai.grainger.uiuc.edu/registry/>
- Et un classement annuel, [Ranking Web of World Repositories](#)

15/02/2011

C. Morel-Pair



15/02/2011

Rechercher | Accueil | Soutenir | Président | Soutenir tout | Respecter la cause | Please non troupée

Geoname Location (1715)

- Africa (35)
  - Botswana (1)
  - South (4)
  - Ethiopia (1)
  - Ghana (2)
  - Kenya (1)
  - Mozambique (1)
  - Namibia (1)
  - Nigeria (2)
  - South Africa (20)
  - Uganda (1)
  - Zimbabwe (1)
- Asia (297)
  - Armenia (1)
  - Azerbaijan (2)
  - Bangladesh (2)
  - China (12)
  - North Korea (4)
  - India (55)
  - Indonesia (17)

## ROAR,

## browse:

country, type,  
software

## s Repositories (ROAR)

Browse by Country | Browse by Type | Browse by Software

Browse by Repository

Please select a value to browse from the list below.

- Repository Software (1979)
  - ARNO (6)
  - BePress (110)
  - CDS Invenio (16)
  - ContentDM by QCLC (5)
  - DIGIRIR (1)
  - DigiTool (10)
  - DIVA (25)
  - DoKS (5)
  - DSpace (802)
  - EDOC (2)
  - EPrints (374)
  - Equella (2)
  - ETD-db (30)
  - Fedora (37)
    - Fez (10)
  - Greenstone (14)
  - HAL (12)
  - i-Tor (1)
  - Intralibrary (1)
  - MITOS (4)
  - MyCoRe (6)
  - Open Journal System (10)
  - Open Repository (17)
  - OPUS (Open Publications System) (42)
  - Other softwares (various) (442)
  - PMB Services (1)
  - SBCAT (1)
  - SciX (3)

## Access Repositories (ROAR)

Content | Browse by Country | Browse by Type | Browse by Software | Register

History

Browse by Repository 1

Please select a value to browse from the list below.

- Database/A&I Index (45)
- Demonstration (19)
- Learning and Teaching Objects (28)
- Other (248)
- Research Cross-Institutional (178)
- Research Institutional or Departmental (1299)
- e-Journal/Publication (113)
- e-Theses (181)

C. Morel-Pair

Copyright 2005-2010 Tim Brody <tdb2 AT eds.soton.ac.uk>, University of Southampton  
This work is licensed under a [Creative Commons License](#).

Copyright 2005-2010 Tim Brody <tdb2 AT eds.soton.ac.uk>, University of Southampton  
This work is licensed under a [Creative Commons License](#).

# Ranking Web of World Repositories

January 10

**About the Ranking**

Top 400

Top 400 Institutional

Directory

Methodology

Best Practices

**Information**

Glossary

Blog

Links

Contact Us

Site Map

Disclaimer

Search Repositories for:

Submit

Category:

All

**CybermetricsLab**

[www.mavir.net](http://www.mavir.net)

WORLD RANK	REPOSITORY		POSITION			
			COUNTRY	SIZE	VISIBILITY	RICH FILES SCHOLAR
1	Arxiv.org e-Print Archive		4	2	2	3
2	CiteSeerX		1	1	411	2
3	Scientific and Technical Information Network		7	7	1	4
4	Social Science Research Network		6	3	51	6
5	Research Papers in Economics		2	6	140	5
6	Smithsonian/NASA Astrophysics Data System		3	4	539	1
7	Hal CNRS		18	9	5	8
8	Institut National de Recherche en Informatique et en Automatique Archive Ouverte		19	11	26	14
9	CERN Document Server		11	16	40	11
10	MIT Dspace		14	34	7	9
11	Ecole Polytechnique Federale de Lausanne Infoscience		8	14	49	95
12	Agecon Search Research in Agricultural and Applied Economics		21	45	4	17
13	University of Michigan Deep Blue		27	28	13	20
14	Hyper Article en Ligne Sciences de l'Homme et de la Société		83	12	122	30
15	Igitur Archive Universiteit Utrecht		26	60	3	28
16	Munich Personal Repec Archive		47	30	14	22
17	Universitat Stuttgart elektronische Hochschulschriften		35	13	37	241
18	University of Southampton: Department of Electronics and Computer Science		63	20	32	67
19	Dissertations of the University of Groningen		40	52	6	33
20	Universidade do Minho Repositorium		20	48	27	40
21	INIST I-Reviews		12	56	45	31
22	Queensland University of Technology ePrints		17	83	21	16

Distingue en 2011 les "archives ouvertes", institutionnelles ou thématiques, et les autres projets : "portails"

**WEBOMETRICS RANK**

VISIBILITY (external inlinks) 50%	SIZE (web pages) 20%
RICH FILES 15%	SCHOLAR 15%



## OAI-PMH

“le pouvoir de la simplicité”

### 1 – archives ouvertes scientifiques

- 1400-1500 archives ouvertes, 15-20 millions de références ...
  - Voie d'or + verte : 15-20 % des publications sont en libre accès
  - Et 85 % des références pointent sur des ressources en libre accès (fr)
- Création de réseaux nationaux et internationaux - projets agrégeant des entrepôts
  - DRIVER : infrastructure digitale pour la recherche en Europe
  - CLARIN : fédérer les données de la recherche en linguistique
- ... Reste des questions sémantiques – étude DRIVER :
  - Usage des éléments de métadonnées et des “sets” OAI
  - Valeurs très hétérogènes : multi-linguisme, mots-clés, structure des dates, auteurs et affiliations ... ..

15/02/2011

C. Morel-Pair

## OAI-PMH, métadonnées et données de la recherche



- CLARIN
  - Projet européen
  - 2011-01 : 25 centres
- Virtual Language Observatory
  - Moissonnage OAI des centres participants
  - Métadonnées
    - DC simple
    - OLAC, profil d'application DC
    - IsoCat, spécifique description fine des corpus langagiers

15/02/2011

C. Morel-Pair

## DRIVER : exemples de valeurs trouvées dans dc:date

September 29–October 28, 51 AD; 1970  
second half of IXth century AD; 1978

**Rebuilt 1984**

Possibly Vth/VIth century AD; 1935

**Planted 1985**

n/a

n.d.

Mid IIInd century AD; 1973

Jul-51

circa 900 AD

ca. 701 BC

Begun 14th century

184-?

18-?

August 23, 2000

between 1827 and 183

XVIII Dynasty

Winter 2003

VIIIth/IXth century AD ? (TC);1965

Vth-VIth century AD (McNamee); IVth century AD (Cribiore); 1982

**era of redevelopment**

**Various (... et : Various)**

2002-00

1980, refurbished 1997

China: Neolithic Period (5000 BCE-ca 1600 BCE)?

19691968

21. Nouemb. Anno. 1564.

**And finissed on the euen of thanunciacion of our  
said bilissid Lady falling on the  
wednesday the xxiiij daye of Marche. in the xix yeer  
of Kyng Edward the fourthe**

[1479]

19193

xxxx Oct xx

1938-05-38

1963 to 1953

[not after 1579]

...

15/02/2011

C. Morel-Pair

## OAI-PMH : le pouvoir de la simplicité 2 – Revues scientifiques

- Des entrepôts OAI-PMH pour signaler les **revues scientifiques** en libre accès
  - BioMed Central, Revues.org, Persée, I-Revues ...
    - Revues nouvelles ou non (rétrounumérisation)
    - Embargo ou non
  - Et certaines plateformes de revues plus commerciales, exemple : CAIRN

15/02/2011

C. Morel-Pair

## Portail SHS ISIDORE



Projet du TGE Adonis Signale et mutualise publications et données de recherche en SHS déc. 2010

Constitué par moissonnage d'entrepôts OAI-PMH + RSS et RDFa + SRU

Interface à facettes Vocabulaires contrôlés

métadonnées exposées en triples RDF

<http://www.rechercheisidore.fr/index>

15/02/2011 C. Morel-Pair



## OAI-PMH le pouvoir de la simplicité - 3

### ■ Des portails et catalogues variés

- Patrimoine-culture, administration, autres
- Interopérabilité de ressources et collections numériques **diverses** via moissonnage OAI-PMH de métadonnées standard
  - + variantes : métadonnées dans tableaux, RSS, soumission ...

### ■ Intégration des enregistrements OAI à des catalogues plus commerciaux

- Scirus, Scopus (Elsevier), EBSCO ... – contractualisation ?
- Rachat d'OAIster par OCLC → Worldcat

### ■ Un moyen simple pour échanger des métadonnées

- Intégration de catalogues d'éditeurs privés exposés en OAI (ONIX souvent) par partenaires ciblés (BNF, INIST ...) → transformations → intégration BD et catalogue → visibilité

15/02/2011

C. Morel-Pair




# OAI-PMH pour la mutualisation des services patrimoine-culture

- Des services en ligne riches, multiformes et éparpillés
  - Musées, archives, bibliothèques ... métadonnées variées !
- Des objectifs
  - Visibilité des sites, recherches transversales
  - Créer des services spécifiques : thématiques, régionaux, événements ...
  - Respect des logiques des producteurs
- ... Des portails et catalogues créés par moissonnage OAI-PMH
  - Mutualisation des fonds numérisés patrimoine-culture, France
    - Via entrepôts Dublin Core, MODS, EAD, DIDL ...
      - Catalogue collectif, portails régionaux, projet européen Michael-Minerva
    - Gallica (Dublin Core) intégré dans Europeana ...

15/02/2011

C. Morel-Pair

## Portails patrimoine






<http://www.numerique.culture.fr/mpf/pub-fr/index.html>

[http://www.culture.fr/fr/sections/collections\\_moteur](http://www.culture.fr/fr/sections/collections_moteur)

<http://bnsa.patrimoines.aquitaine.fr/>

15/02/2011



## Les entrepôts OAI de la Bibliothèque nationale de France

Pour faciliter l'accès à ses documents numérisés et aux métadonnées descriptives des documents conservés dans ses collections, la Bibliothèque nationale de France implémente le protocole **OAI-PMH**.

[ [bibnum.bnf.fr](http://bibnum.bnf.fr) ]

Deux entrepôts OAI sont actuellement en ligne : OAI-NUM et OAI-CAT. Ils sont organisés en ensembles ou sets regroupant les notices :

- par types de documents,
- par dossiers thématiques ou par fonds correspondant à l'organisation des collections,
- par indexation sujet suivant la classification décimale Dewey.

Les principes retenus à la BnF dans l'utilisation des éléments de métadonnées du Dublin Core non qualifié sont présentés dans le **Guide d'utilisation du Dublin Core à la BnF (septembre 2008)** (16 pages - pdf - 184Ko).

### OAI-NUM

L'entrepôt OAI-NUM signale les documents provenant des collections numériques. Il est enrichi toutes les semaines de milliers de notices, dans le cadre de la **politique de numérisation de la BnF**.

Il contient actuellement 340 000 notices donnant accès à des documents consultables dans la bibliothèque numérique **Gallica**, dont la nouvelle interface est actuellement accessible en version bêta sous le nom de **Gallica2**. Ces 340 000 notices correspondent à environ 72 000 monographies, 1 700 titres de périodiques (les 1700 titres et les 245 000 fascicules sont décrits dans deux sets distincts, **gallica:periodiques:titres** et **gallica:periodiques:fascicules**), 80 000 images, 3 700 cartes et documents géographiques.

Par ailleurs, les notices des manuscrits numérisés et des documents sonores seront versées dans l'entrepôt d'ici la fin de l'année 2008. Les collections de **Mandragore** (base iconographique du département des Manuscrits) et de la **Banque d'images** (département de la Reproduction) viendront également enrichir cet ensemble.

Les notices sont disponibles sous deux formats : Dublin Core simple et MPEG21.

Dans le cadre de la coopération avec **The European Library**, la BnF enrichira l'entrepôt avec le **profil d'application TEL**.

L'URL donnant accès à la page de description de OAI-NUM est : <http://oai.bnf.fr/oai2/OAIHandler?verb=Identify>. L'URL de base est : <http://oai.bnf.fr/oai2/OAIHandler>.

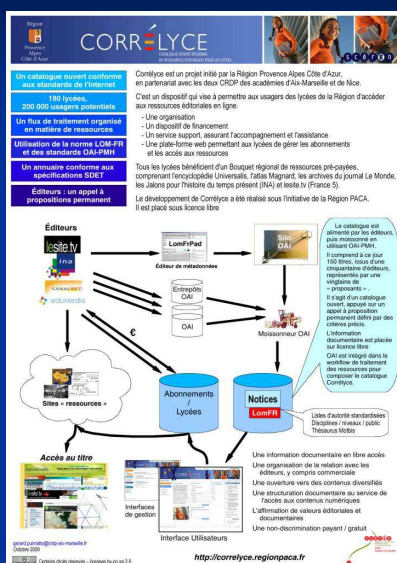
### OAI-CAT

L'entrepôt OAI-CAT a vocation à contenir les notices décrivant tous les documents conservés à la BnF, qu'ils soient numérisés ou non. Il a été mis en ligne avec un premier versement totalisant plus de 9,8 millions de notices bibliographiques provenant du catalogue **BN-Opale plus**. Il sera également alimenté ultérieurement par les informations descriptives issues du catalogue **BnF archives et manuscrits**.

L'URL donnant accès à la page de description de OAI-CAT est : <http://catoai.bnf.fr/oai2/OAIHandler?verb=Identify>. L'URL de base est <http://catoai.bnf.fr/oai2/OAIHandler>.

→ Europeana et autres catalogues

## Autres projets ...



The screenshot shows the homepage of the Portail du Patrimoine Oral. It features a large image of a clock face with a hand pointing to the number 12. The text "Portail du Patrimoine Oral" is prominently displayed. Below the image, there is a section titled "Pourquoi un Portail du Patrimoine Oral (PPO) ?" which explains the project's goals. It mentions the need for a central platform for oral heritage, the importance of digital preservation, and the goal of making this heritage accessible to all. The text also mentions the project's funding by the French Ministry of Culture and the French Republic. Below this, there is a section titled "Identité :" which provides information about the project's identity, including its name, acronym, and contact details. The bottom of the page features a navigation bar with links to various sections of the portal.

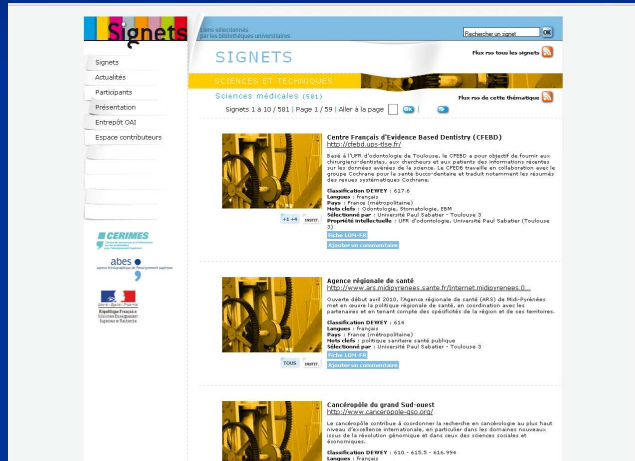
<http://www.portaildupatrimoineoral.org/>

Source : FReDOC 2009,  
<http://renatis.cnrs.fr/spip.php?article156>

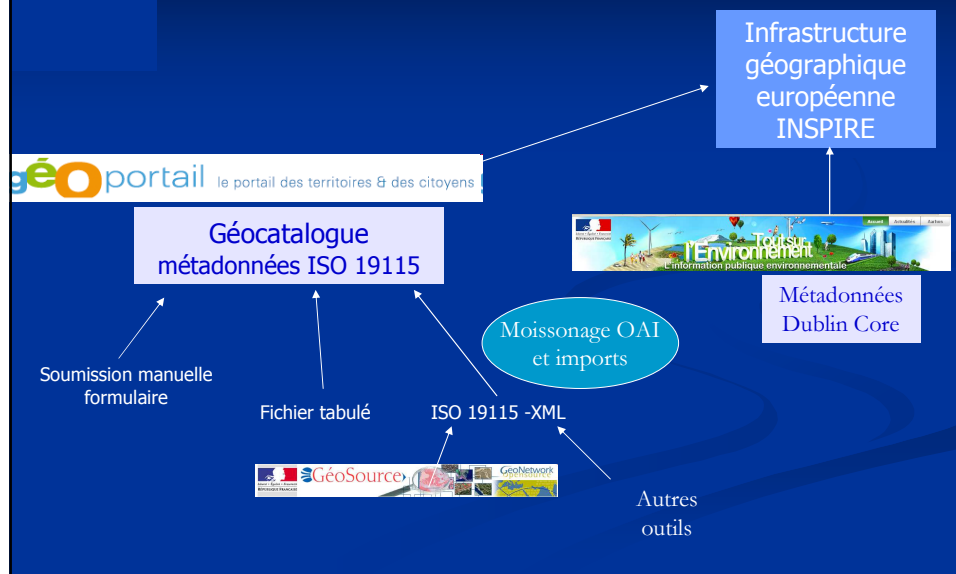
C. Morel-Pair

15/02/2011

... Les Signets des BU exposent leurs métadonnées dans un entrepôt OAI-PMH



## Portails information publique

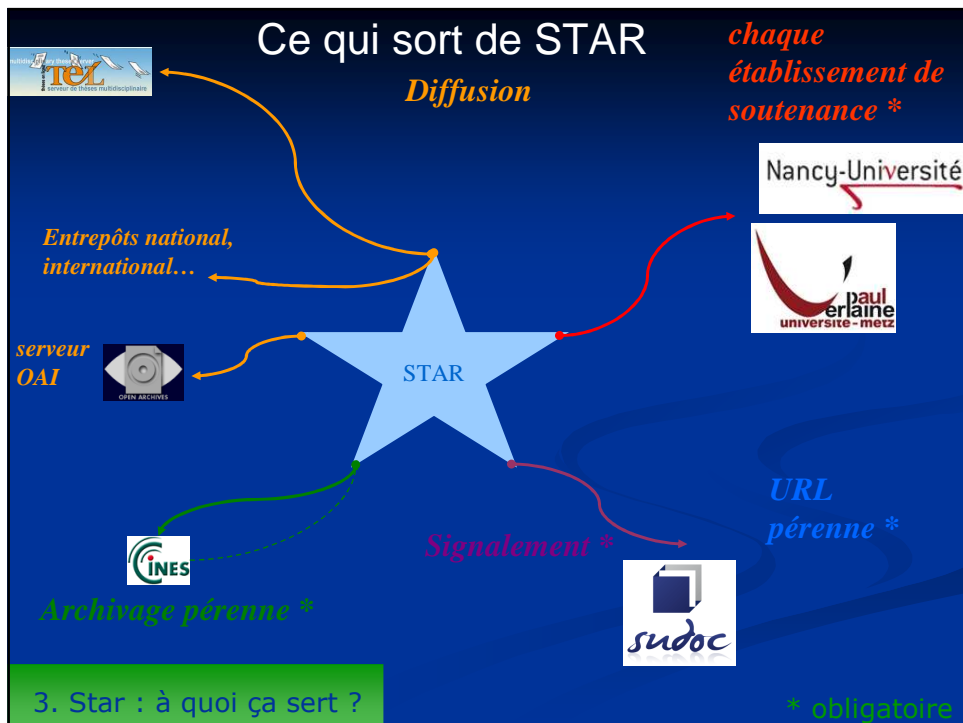




## Ce qui rentre dans STAR



## Ce qui sort de STAR



## Derrière STAR : métadonnées TEF

- Un “profil d’application” **Dublin Core + METS**
  - Version 1 (2005) : métadonnées surtout descriptives
    - **Dublin Core** qualifié + **MODS et ETD-MS**
  - Version 2 (2006) : une seule description pour toutes les manifestations de la thèse, et métadonnées de gestion
    - encapsulation **METS** + droits **METS**
  - Valeurs contrôlées : vocabulaires, autorités MADS...
- Source diapos précédentes : Isabelle Mauger Perez – ABES, in Journée d’étude sur les thèses électroniques, 26 mars 2009, Médial, Nancy

15/02/2011

C. Morel-Pair



## 3 - ORI-OAI, Outil de Référencement et d’Indexation - Réseau de portails OAI

- Travaux de plusieurs établissements représentatifs enseignement supérieur, 2001-2008
- Mise en place d'un système ouvert pour
  - **gérer et publier** tous les documents numériques produits par les établissements universitaires
    - Pédagogiques, travaux de recherche, administratifs
  - les **partager** avec d'autres établissements
  - les **valoriser** par une indexation de qualité
  - les **rendre accessibles** à distance, selon les droits définis, dans des interfaces ergonomiques

15/02/2011

C. Morel-Pair



## ORI-OAI principes d'un système

- Un processus de gestion et publication des ressources numériques
  - Un référentiel unique de ressources connecté au SI
  - Un work-flow impliquant auteur, avec processus validation-suivi paramétrable
  - Web, droits d'accès, métadonnées standard
  - Archivage pérenne
- Partage avec communauté(s)
  - Réseau de portails via **OAI-PMH**
  - Interopérabilité avec autres "systèmes" : ENT, UN, HAL, STAR, autorités ABES, CEPIA ...
- Open source - simplicité

15/02/2011

C. Morel-Pair



## ORI-OAI, choix techniques

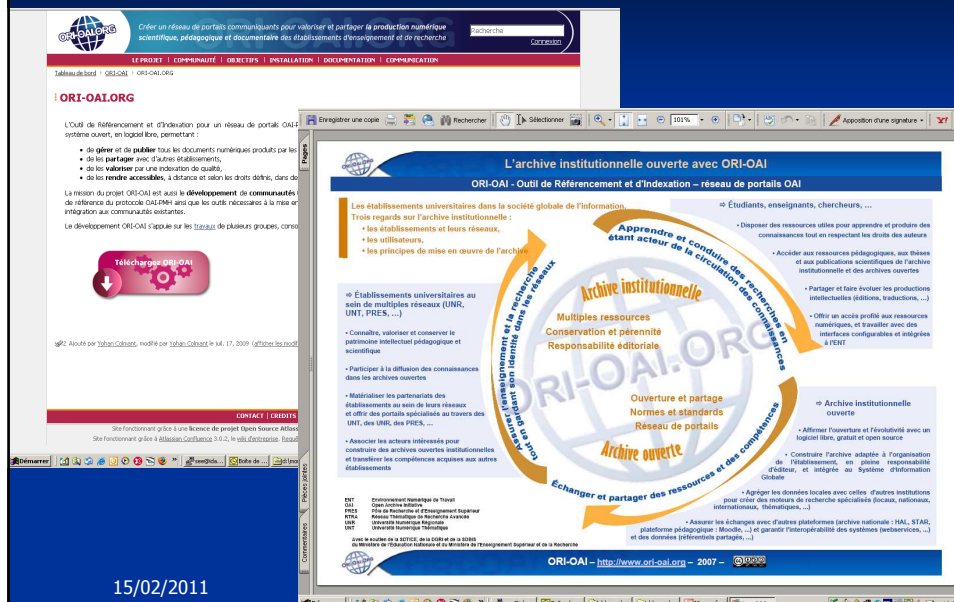
- Des métadonnées interopérables
    - **XMI**, **Dublin Core**, **LOM** et **LOM-fr**, **TEF** ...
    - Classifications paramétrables articulées sur des standards comme Dewey
  - Des applications existantes
    - OAI-PMH, JSR 168, Shibboleth ...
  - Des outils libres
- 7 modules communiquants ... Version 1

15/02/2011

C. Morel-Pair



# Réseau de portails Enseignement Supérieur et archive institutionnelle



## ORI-OAI en 2011

- **Nombreux projets d'archive institutionnelle dans les universités**
- **Bordeaux 1 : d'une plateforme de diffusion des thèses ...**
  - Un historique - choix des outils ORI - thèse numérique obligatoire 2008
    - Simplicité : export des notices TEF de STAR, transformation automatisée → Dublin Core → intégration
    - En cours : gestion des droits
  - ... à une archive institutionnelle (?)
  - Intégrant ressources pédagogiques et publications
    - Moissonnage des entrepôts des UNT + (prévu) de HAL
    - Réflexions: ergonomie recherche et dépôts ...
- **ORI-OAI V2 : vers une meilleure interopérabilité**
  - Permettre le dépôt local puis l'export des données vers STAR, HAL

15/02/2011

C. Morel-Pair

- Métadonnées ? Généralités et standards
- L'interopérabilité en action
- Les standards en action : outils libres et propriétaires



15/02/2011

C. Morel-Pair



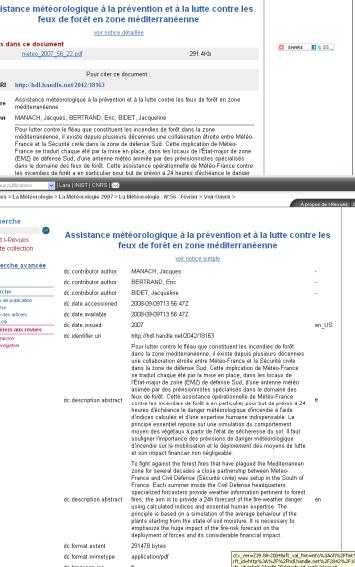
## Plateformes de publication, métadonnées et OAI-PMH

- Exemple : DSpace (MIT)
  - À l'INIST : I-Revues, <http://irevues.inist.fr> et LARA, <http://lara.inist.fr>
- Organisation arborescente - navigation et recherche
- Ressources en accès libre ou non
  - Paramétrage des accès
- Services : alertes et RSS, widgets web 2.0, statistiques
- Exposition OAI-PMH : DC simple par défaut
- Souci de la pérennité
  - Attribution d'un URI (handle) par item
  - Formats des ressources, signature numérique

15/02/2011

C. Morel-Pair

# DSpace et I-revues



- **Métadonnées internes DC**  
**qualifié + METS (collections)**
  - Affichage paramétrable
  - autres paramétrables
- **Exposition OAI-PMH**
  - **DC simple** par défaut
- **Acquisition des données par**
  - Formulaire de soumission
  - Workflow paramétrable
  - Imports en masse :  
ressources accompagnée de leurs métadonnées  
Dublin Core en XML (ou autres)
- **Source : transformation de l'existant**

[illegible]

- Recherche
- Revue & Congrès
- Date de publication
- Auteur
- Titre des articles
- Mot-clé
- Abonné(e)s aux revues
- Déconnexion
- Modifier mon profil
- Soumissions

Nom du champ:	dc. <input type="text"/> - <input type="text"/>
Note d'application:	<input type="text"/>
Notes additionnelles concernant ce champ de métadonnées.	
<input type="button" value="Ajouter"/>	

Identifiant	Champ	Note d'application
2	<a href="#">dc.contributor.advisor</a>	Use primarily for their advisors.
69	<a href="#">dc.contributor.affiliation</a>	Affiliation, conforme recommandations DCM, juin 05. Remis en usage en 2010 avec bascule de organisation affiliation
3	<a href="#">dc.contributor.author</a>	OAI-4I le passe en créateur conformément au schéma OAI-PMH
4	<a href="#">dc.contributor.editor</a>	
5	<a href="#">dc.contributor.illustrator</a>	
6	<a href="#">dc.contributor.other</a>	
70	<a href="#">dc.contributor.partner</a>	Crée pour l'OMNT, pour entrer les organismes clés dans le texte et ayant participé au projet
1	<a href="#">dc.contributor</a>	A person, organization, or service responsible for the content of the resource. Catch-all for unspecified contributors.
7	<a href="#">dc.coverage.spatial</a>	Spatial characteristics of content.
8	<a href="#">dc.coverage.temporal</a>	Temporal characteristics of content.
9	<a href="#">dc.creator</a>	Do not use; only for harvested metadata. ne pas l'utiliser ici car on n'a plus droit qu'à un seul en saisie man!
11	<a href="#">dc.date.accessioned</a>	Date DSpace takes possession of item.
12	<a href="#">dc.date.available</a>	Date or date range item becomes available to the public.
13	<a href="#">dc.date.copyright</a>	Date of copyright.
14	<a href="#">dc.date.created</a>	Date of creation or manufacture of intellectual content if different from date issued.
15	<a href="#">dc.date.issued</a>	Date of publication or distribution.
16	<a href="#">dc.date.submitted</a>	Recommended for theses/dissertations.
10	<a href="#">dc.date</a>	Use qualified form if possible.

- Nombreux formats de fichiers texte, image, vidéo
  - Paramétrable
  - Suivis, connus ou non de la plateforme

Cette liste des formats de fichiers fournit des informations sur les formats connus et leur niveau de support. Avec cet outil, vous pouvez modifier ou ajouter de nouveaux formats. Les formats déclarés internes sont cachés du point de vue de l'utilisateur, et ne sont disponibles qu'à des fins d'administration.

Identifiant	Nom	Type MIME	Niveau de support
1	Unknown	image/gif (inconnu)	Support
2	Licence	text/plain (inconnu)	Contenu
3	Adobe PDF	application/pdf	Contenu
4	XML	text/xml	Contenu
5	Text	text/plain	Contenu
6	HTML	text/html	Contenu
7	Microsoft Word	application/msword	Contenu
8	Microsoft PowerPoint	application/vnd.ms-powerpoint	Contenu
9	Microsoft Excel	application/vnd.ms-excel	Contenu
10	MARC	application/marc	Contenu
11	JPEG	image/jpeg	Contenu
12	GIF	image/gif	Contenu
13	image/png	image/png	Contenu
14	TIF	image/tif	Contenu
15	AIFF	audio/aiff	Contenu
16	audio/basic	audio/basic	Contenu
17	WAV	audio/wav	Contenu
18	MPEG	video/mpeg	Contenu
19	RTE	text/richext	Contenu
20	Microsoft Visio	application/vnd.ms-visio	Contenu
21	FMPI	application/flexmaker	Contenu
22	BMP	image/ms-bmp	Contenu
23	Photoshop	application/photoshop	Contenu
24	Postscript	application/postscript	Contenu
25	Video Quicktime	video/quicktime	Contenu
26	MPEG Audio	audio/mpeg	Contenu
27	Microsoft Project	application/vnd.ms-project	Contenu
28	Mathematica	application/mathematica	Contenu
29	Latex	application/latex	Contenu
30	TeX	application/latex	Contenu
31	TeX.dvi	application/dvi	Contenu
32	SGML	application/sgml	Contenu
33	WordPerfect	application/wordperfect.1	Contenu
34	RealAudio	audio/x-realaudio	Contenu



## et Dublin Core

- Plate-forme éditoriale libre développée par le CLEO
  - Revues SHS, présentation paramétrable
    - Exemple : [www.revues.org](http://www.revues.org)
- Work-flow de soumission
  - D'un document-auteur Word ou Open Office à la mise en ligne en format **XHTML**
  - **Document** et **métadonnées**, avec index de recherche
  - En amont : nécessiter de styler les documents selon le modèle .dot Lodel + identification et niveaux d'autorisation

15/02/2011

C. Morel-Pair

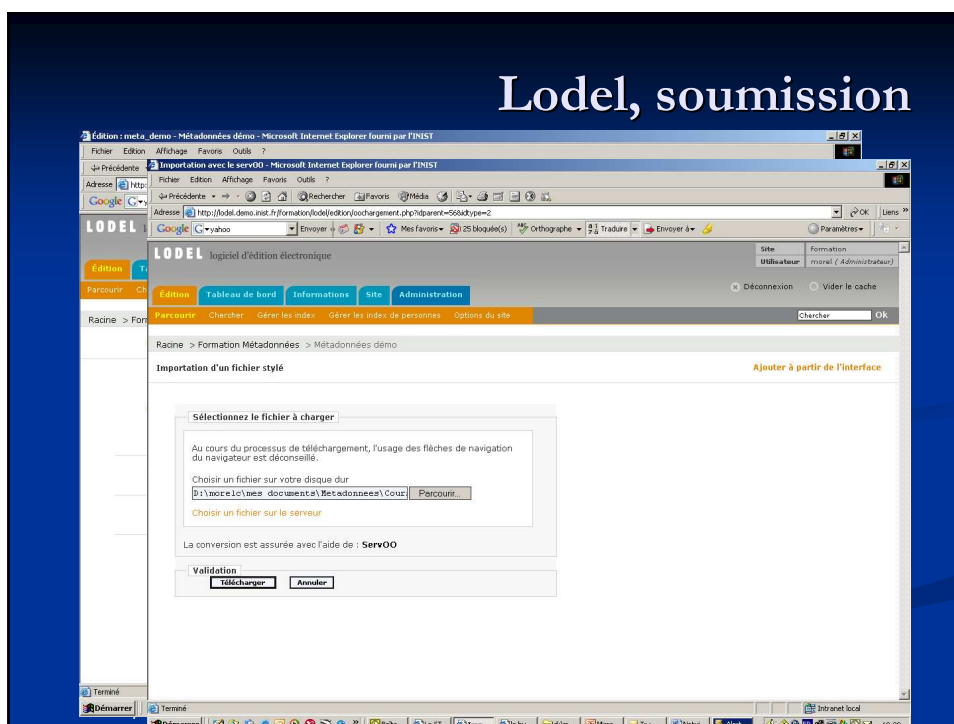
## D'un article publié ailleurs à la publication avec Lodel ...

(Exemple pour TP)

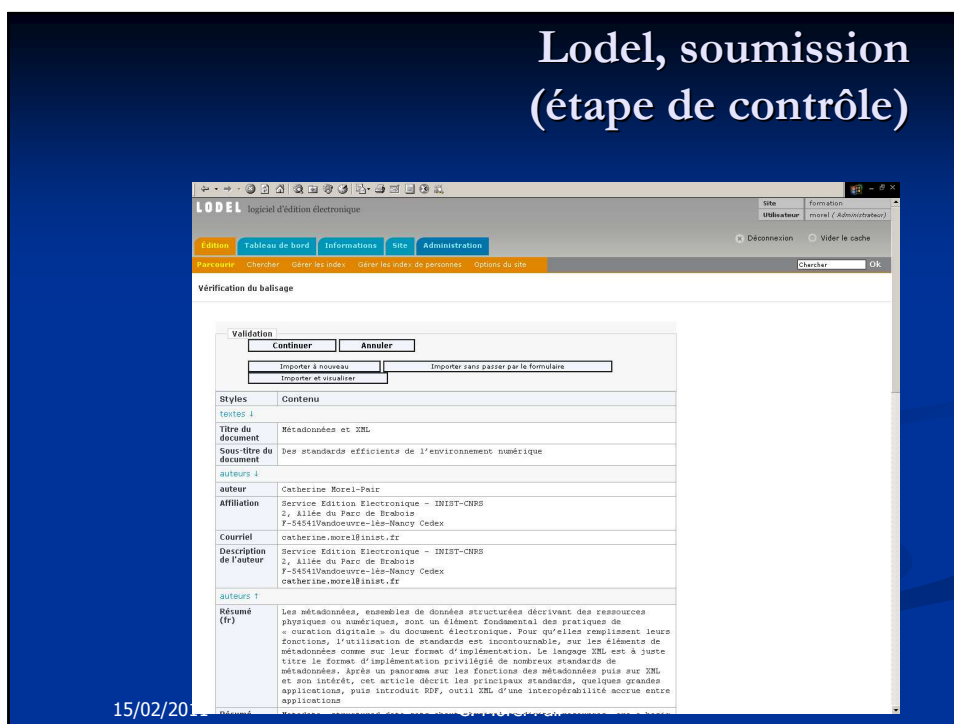
The screenshot displays a Microsoft Word document titled "Métadonnées-et-XML". The document content includes a title "Des standards efficaces de l'environnement numérique", an author "Catherine Morel-Pair", and a detailed abstract in French. The abstract discusses the importance of metadata standards in digital environments and mentions the use of XML and RDF. The document is shown in a window titled "Meta\_CMO\_monstyle.doc - Microsoft Word".

15/02/2011

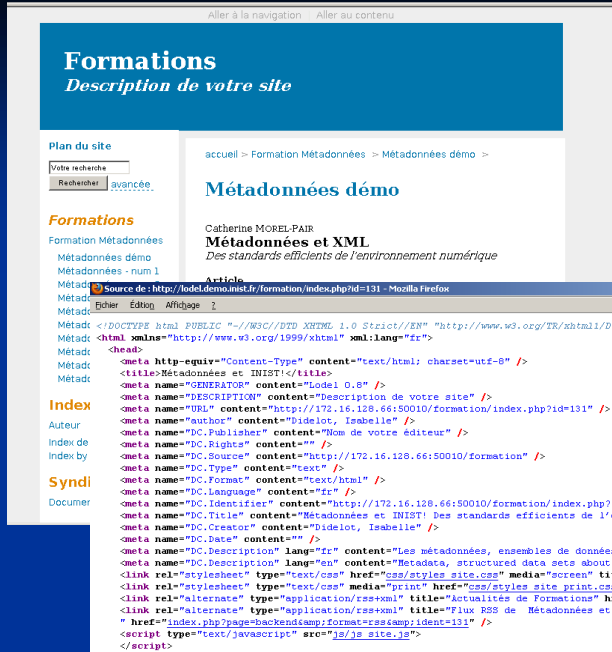
# Lodel, soumission



# Lodel, soumission (étape de contrôle)



15/02/2011

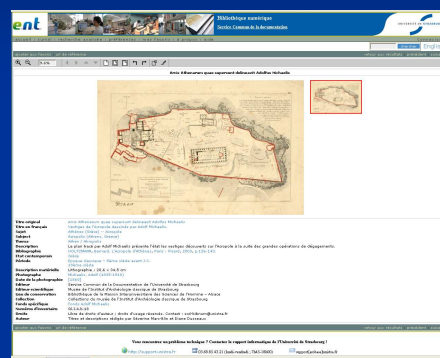


# Résultat

Lodel génère aussi du DC en XML pour son entrepôt OAI-PMH

## Logiciels propriétaires pour les bibliothèques numériques – 2 exemples

- **CONTENTdm** (OCLC)
  - XML, DC, VRA, METS/ALTO, RSS
  - vocabulaires contrôlés variés
  - exposition OAI-PMH
- **Divvalib** (Pollinum, BNSA)
  - “Bibliothèques numériques full-media de nouvelle génération”
  - fonctions riches de publication, recherche, navigation et affichage, collaboration
  - Enrichissement et conversions des documents
  - XML, RDF, METS, ALTO, EAD, DC, ONIX
  - OAI-PMH



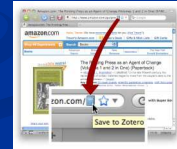
contentDM : [fonds numérisés de l'Université de Strasbourg](#),



# Logiciels documentaires

## ... SIGB ... GED ...

- Dans beaucoup de logiciels aujourd'hui
  - Imports et exports XML, éléments variés ; annonce nouveautés RSS
  - Exposition et moissonnage OAI-PMH
- ... Et plus (Koha, SIM ... )
  - couches internes XML encodant Dublin Core + (selon les cas) EAD, MODS, METS ...
- Outils Web 2 de gestion de références bibliographiques et bibliothèques
  - Zotero “reconnait” → importe, et exporte : Dublin Core, MODS, FOAF ... en XML et RDF CoiNs ...
  - Connotea : imports et exports MODS, RDF ...



15/02/2011

C. Morel-Pair



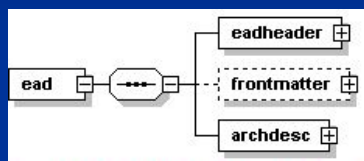
- Standardiser et mettre en ligne les inventaires archivistiques, avec des fonctions de recherche et de navigation → “instruments de recherche”
- ... Dans un format XML pour permettre
  - Pérennisation : conservation et migrations
  - Accessibilité accrue
    - Présentations diverses adaptées au contexte d'édition - public cible
    - Fonctions de recherche et navigation
  - Echanges et mutualisation de données standardisées
    - Vastes réservoirs archivistiques complémentaires
  - Transformations vers d'autres applications : catalogues de bibliothèques, OAI-PMH ...

# EAD

## “Le” standard archivistique

- Basé sur la norme ISAD(G), fonctions et structure des instruments de recherche
  - Respect des fonds, de leur contexte de constitution et de leur structuration
    - Un enregistrement par “unité documentaire”
    - Logique de production → composants physiques hiérarchisés
    - Description du général au particulier - héritage
      - fonds/collection/série/“boîte”
- Maintenu par la Bibliothèque du Congrès

## EAD, structure



- **<eadheader>** métadonnées décrivant le **document EAD**
- **<archdesc>** description structurée du fonds

# EAD, archDesc, vue simplifiée ...

**archDesc**

- did** (solid box)
  - Identifiant, titre, origine, date, container, abstract, physloc, **dao** ...
- bioghist (dashed box)
- scopecontent (dashed box)
- controllaccess (dashed box)
- ... .. (dashed box)
- dsc** (dashed box)
  - did** (solid box)
    - Outils dédiés : **Pleade**, Archivistic Toolkit ...
  - c** (dashed box)
    - dsc** (dashed box)
      - c** (dashed box)
      - c** (dashed box)
  - c** (dashed box)

**Etc ...**

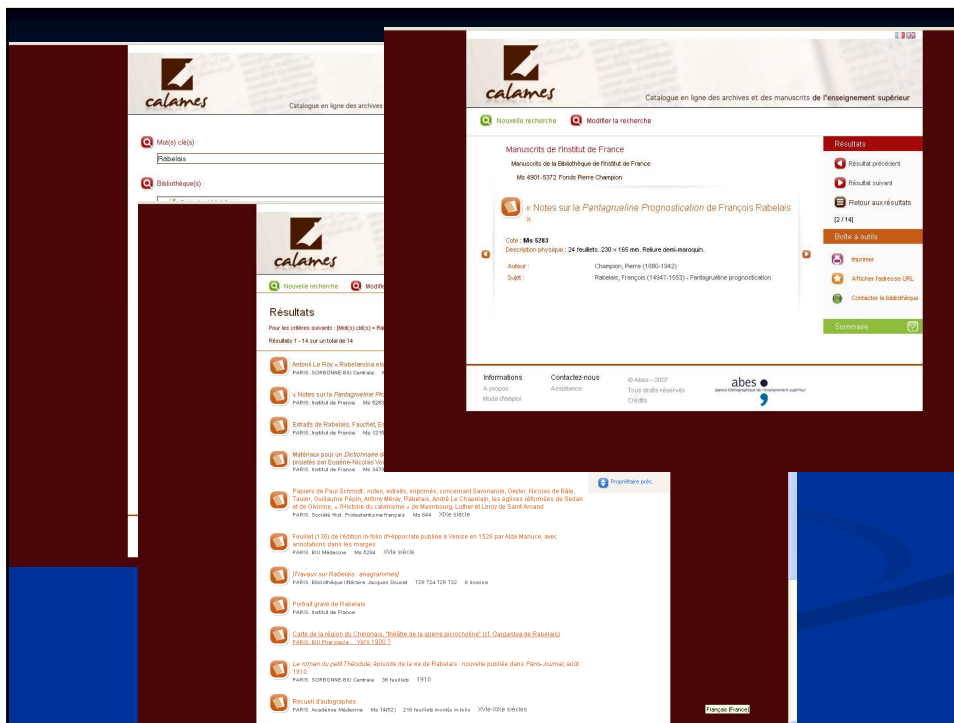
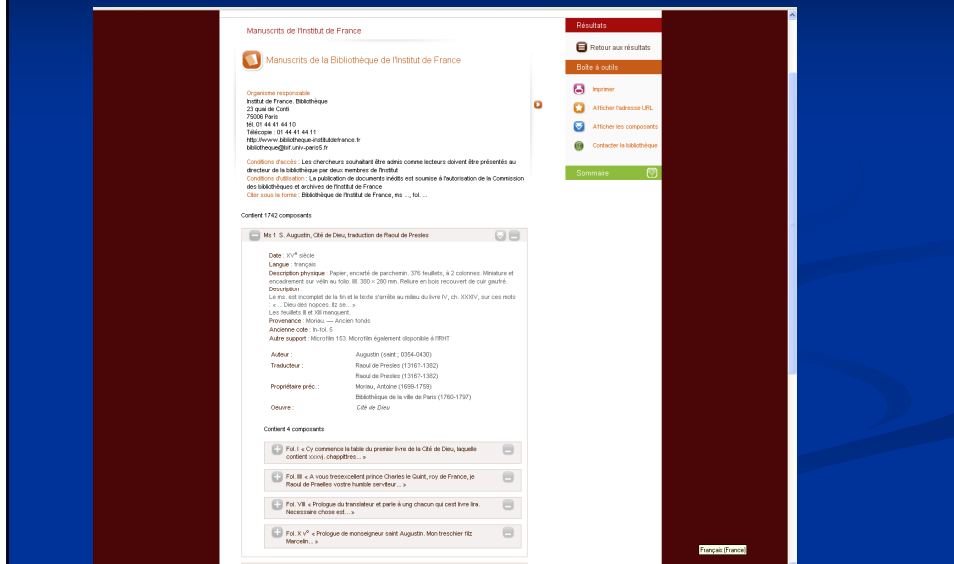
# EAD

## ... bien présent sur le Web

- Conversion d'inventaires archivistiques papier
  - Centres d'archives nationales, régionales, municipales ... tous pays
    - Ex : [Genève](#)... pour la possibilité de voir le fichier EAD brut
- Catalogues de collections spécialisées de bibliothèques
- Mutualisation
  - [Calames](#), catalogue collectif des archives et manuscrits de l'enseignement supérieur, ABES-BNF
  - Transformation des signalements disponibles
- **Premières** descriptions de **collections**

# Calames

## Un des organismes participant



# A suivre ... à suivre ... à suivre

## Le web sémantique

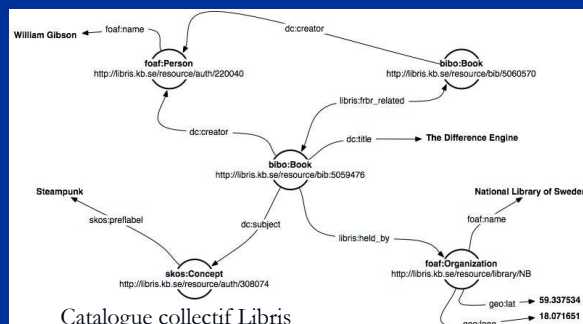
- Un web de données structurées, ouvertes, liées, interprétables par les machines → nouvelles données
  - Un cadre d'écriture : RDF
    - Donnée = triples "Sujet prédicat objet" sur "tout"
    - Plus de 10 milliards de triples communiquent sur le web !
  - Des ontologies (... en RDF) pour renseigner les triples
    - Personnes (FOAF, vCard, VIAF, SIOC)
    - Connaissances scientifiques (Bioportal : ontologie en biologie) ... pratiques métiers ...
    - Événements (iCalendar) ou lieux (geo, geonames, INSEE ...)
    - Ressources (DC, Bibo) et droits (CC)
    - Concepts (RAMEAU, LCSH, Eurovoc ... en SKOS, ou plus) ...
  - Des URI, identifiants pérennes et "déréférencables" pour toutes ces "choses" ...
    - ... web de métadonnées !

15/02/2011

C. Morel-Pair

# A suivre ... à suivre ...

- Modéliser et ouvrir nos catalogues sur le web de données ?
  - Qu'est-ce qu'un catalogue ? Les tendances actuelles : un nouveau modèle de données, les FRBR, le projet de code international de catalogage RDA, le Web sémantique, l'interopérabilité, le Web de données, les perspectives, E. Bermès, conférence URFIST-CFCB Bordeaux, 6 novembre 2010



15/02/2011

C. Morel-Pair

C'est parti en France aussi ...  
à l'ABES et à la BNF

Un groupe W3C "bibliothèques  
et web de données"